STIC Search Report
Biotech-Chem Library

STIC Database Tracking Number: 169374

TO: Bao-Qun Li **Location:** 3C18 **Art Unit: 1648** **Thursday, November 03, 2005** **Case Serial Number: 09/980559**	**From: Edward Hart** **Location: Biotech-Chem Library** **REM-1A55** **Phone: 571-272-2512** **edward.hart@uspto.gov**

Search Notes

Examiner Li,

Here are the results of the search you requested.

Please feel free to contact me if you have any questions.

Edward Hart

This Page Blank (uspto)

169374

STIC-Biotech/ChemLib

From: Li, Bao-Qun
Sent: Monday, October 24, 2005 1:51 PM
To: STIC-Biotech/ChemLib
Subject: 09/980,559

Please do the nucleic acid sequence homology and interference searches against amino acid sequence of SEQ ID NO: 2 and nucleic acid sequence of SEQ ID NO: 1, especialy from nucleotide 341-9439 of SEQ ID NO: 1. Thanks.

Bao Qun Li M.D
TC 1600
Art Unit 1648
Tel. 517-272-0904
REM, 3C18
Rm. 3D24

Searcher: Doan
Searcher Phone:
Date Searcher Picked up:
Date completed: 11/3/05
Searcher Prep Time:
Online Time:

Type of Search
NA# 1 AA#: 1
S/L: Oligomer:
Encode/Transl:
Structure #: Text:
Inventor: Litigation:

Vendors and cost where applicable
STN:
DIALOG:
QUESTEL/ORBIT:
LEXIS/NEXIS:
SEQUENCE SYSTEM:
WWW/Internet:
Other (Specify):

This Page Blank (uspto)

STIC SEARCH RESULTS FEEDBACK FORM

Biotech-Chem Library

Questions about the scope or the results of the search? Contact ***the searcher or contact:***

Mary Hale, Information Branch Supervisor
Remsen Bldg. 01 D86
571-272-2507

Voluntary Results Feedback Form

- *I am an examiner in Workgroup:* ☐ *Example: 1610*
- *Relevant prior art **found**, search results used as follows:*
 - ☐ 102 rejection
 - ☐ 103 rejection
 - ☐ Cited as being of interest.
 - ☐ Helped examiner better understand the invention.
 - ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature
 (journal articles, conference proceedings, new product announcements etc.)

- *Relevant prior art **not found:***
 - ☐ Results verified the lack of relevant prior art (helped determine patentability).
 - ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC-Biotech-Chem Library, Remsen Bldg.

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 31, 2005, 14:58:45 ; Search time 24983 Seconds
(without alignments)
17647.761 Million cell updates/sec

Title: US-09-980-559-1_COPY_341_9439
Perfect score: 9099
Sequence: 1 atgagcacaaatcctaaacc.....ttttcctactccccgctcgg 9099

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	9099	100.0	9711	6	AX057086	AX057086 Sequence
2	9099	100.0	9711	6	AX057317	AX057317 Sequence
3	9099	100.0	9711	6	AX057395	AX057395 Sequence
4	9099	100.0	9711	14	AF177036	AF177036 Hepatitis
5	8497.4	93.4	9589	14	HPCPOLP	D00944 Hepatitis C
6	8494.2	93.4	9589	6	E07361	E07361 gRNA of Hep
7	8494.2	93.4	9589	6	E07362	E07362 cDNA of Hep
8	8494.2	93.4	9589	6	I12861	I12861 Sequence 2
9	8025.4	88.2	9659	14	AB047641	AB047641 Hepatitis
10	7895.4	86.8	9416	14	AF238481	AF238481 Hepatitis
11	7893.6	86.8	9653	14	AF169004	AF169004 Hepatitis
12	7878.6	86.6	9700	14	AF169005	AF169005 Hepatitis
13	7869.8	86.5	9416	14	AF238482	AF238482 Hepatitis
14	7858.6	86.4	9661	14	AF169002	AF169002 Hepatitis
15	7852.2	86.3	9416	14	AF238483	AF238483 Hepatitis
16	7841.6	86.2	9416	14	AF238484	AF238484 Hepatitis
17	7836.6	86.1	9683	14	AB047645	AB047645 Hepatitis
18	7827.4	86.0	9693	14	AF169003	AF169003 Hepatitis
19	7810.8	85.8	9689	14	AY746460	AY746460 Hepatitis
20	7809.4	85.8	9416	14	AF238485	AF238485 Hepatitis
21	7801.4	85.7	9691	14	AB047644	AB047644 Hepatitis
22	7768.4	85.4	9674	14	AB047640	AB047640 Hepatitis
23	7726.2	84.9	9678	14	AB047642	AB047642 Hepatitis
24	7652.6	84.1	9641	14	AB047643	AB047643 Hepatitis
25	7513.4	82.6	9678	6	BD160781	BD160781 Gene of f
26	7513.4	82.6	9678	14	AB047639	AB047639 Hepatitis
27	6099.4	67.0	9513	14	D50409	D50409 Hepatitis C
28	6006.2	66.0	9488	14	AB031663	AB031663 Hepatitis
29	5689.4	62.5	9406	14	AY232740	AY232740 Hepatitis
30	5686.2	62.5	9406	14	AY232741	AY232741 Hepatitis
31	5673.4	62.4	9406	14	AY232731	AY232731 Hepatitis
32	5668.6	62.3	9406	14	AY232730	AY232730 Hepatitis
33	5659	62.2	9406	14	AY232749	AY232749 Hepatitis
34	5655.8	62.2	9406	14	AY232745	AY232745 Hepatitis
35	5653.8	62.1	9511	6	E07468	E07468 cDNA sequen
36	5653.8	62.1	9511	6	I12865	I12865 Sequence 7
37	5652.6	62.1	9406	14	AY232747	AY232747 Hepatitis
38	5652.6	62.1	9406	14	AY232748	AY232748 Hepatitis
39	5651	62.1	9406	14	AY232744	AY232744 Hepatitis
40	5649.4	62.1	9406	14	AY232746	AY232746 Hepatitis
41	5649.4	62.1	9511	6	E07467	E07467 gDNA sequen
42	5637.2	62.0	9406	14	AY232742	AY232742 Hepatitis
43	5636.6	61.9	9406	14	AY232733	AY232733 Hepatitis
44	5630.8	61.9	9406	14	AY232743	AY232743 Hepatitis
45	5629.2	61.9	9406	14	AY232735	AY232735 Hepatitis

ALIGNMENTS

RESULT 1
AX057086
LOCUS AX057086 9711 bp DNA linear PAT 17-JAN-2001
DEFINITION Sequence 1 from Patent WO0075338.
ACCESSION AX057086
VERSION AX057086.1 GI:12309919
KEYWORDS .
SOURCE Hepatitis C virus
ORGANISM Hepatitis C virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.
REFERENCE 1
AUTHORS Yanagi,M., Bukh,J., Emerson,S.U. and Purcell,R.H.
TITLE Cloned genone of infectious hepatitis c virus of genotype 2a and uses thereof
JOURNAL Patent: WO 0075338-A 1 14-DEC-2000;
THE GOVERNMENT OF THE UNITED STATES OF AMERICA (US)
FEATURES Location/Qualifiers
source 1. .9711
/organism="Hepatitis C virus"
/mol_type="unassigned DNA"
/db_xref="taxon:11103"
ORIGIN

Query Match 100.0%; Score 9099; DB 6; Length 9711;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 9099; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 400

Qy         61 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 460

Qy        121 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 520

Qy        181 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        521 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 580
Qy        241 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 300
Db        581 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 640
Qy        301 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 360
Db        641 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 700
Qy        361 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 420
Db        701 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 760
Qy        421 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 480
Db        761 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 820
Qy        481 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 540
Db        821 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 880
Qy        541 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 600
Db        881 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 940
Qy        601 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 660
Db        941 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 1000
Qy        661 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 720
Db       1001 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 1060
Qy        721 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 780
Db       1061 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 1120
Qy        781 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 840
Db       1121 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 1180
Qy        841 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 900
Db       1181 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 1240
Qy        901 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 960
Db       1241 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 1300
Qy        961 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1020
Db       1301 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1360
Qy       1021 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1080
Db       1361 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1420
Qy       1081 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1140
Db       1421 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1480
Qy       1141 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1200
Db       1481 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1540
Qy       1201 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1260
Db       1541 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1600
Qy       1261 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1320
Db       1601 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1660
```

```
Qy       1321 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1380
Db       1661 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1720
Qy       1381 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1440
Db       1721 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1780
Qy       1441 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1500
Db       1781 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1840
Qy       1501 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1560
Db       1841 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1900
Qy       1561 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1620
Db       1901 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1960
Qy       1621 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 1680
Db       1961 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 2020
Qy       1681 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 1740
Db       2021 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 2080
Qy       1741 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 1800
Db       2081 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 2140
Qy       1801 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 1860
Db       2141 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 2200
Qy       1861 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 1920
Db       2201 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 2260
Qy       1921 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 1980
Db       2261 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 2320
Qy       1981 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2040
Db       2321 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2380
Qy       2041 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2100
Db       2381 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2440
Qy       2101 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2160
Db       2441 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2500
Qy       2161 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2220
Db       2501 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2560
Qy       2221 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2280
Db       2561 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2620
Qy       2281 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATATTTTGTCATCTTTTTCGTGGCTGCTTGG 2340
Db       2621 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATATTTTGTCATCTTTTTCGTGGCTGCTTGG 2680
Qy       2341 TACATCAAGGGTCGGGTAGTCCCCTTAGCTACCTATTCCCTCACTGGCCTGTGGTCCTTT 2400
Db       2681 TACATCAAGGGTCGGGTAGTCCCCTTAGCTACCTATTCCCTCACTGGCCTGTGGTCCTTT 2740
```

```
Qy    2401 AGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTGCATGGC 2460
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2741 AGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTGCATGGC 2800

Qy    2461 CAGATAGGAGCGGCTCTGCTGGTAATGATCACTCTCTTTACTCTCACCCCCGGGTATAAG 2520
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2801 CAGATAGGAGCGGCTCTGCTGGTAATGATCACTCTCTTTACTCTCACCCCCGGGTATAAG 2860

Qy    2521 ACCCTTCTCAGCCGGTTTTTGTGGTGGTTGTGCTATCTTCTGACCCTGGGGGAAGCTATG 2580
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2861 ACCCTTCTCAGCCGGTTTTTGTGGTGGTTGTGCTATCTTCTGACCCTGGGGGAAGCTATG 2920

Qy    2581 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGCATCATATGGGCC 2640
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2921 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGCATCATATGGGCC 2980

Qy    2641 GTCGCCATATTCTACCCAGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 2700
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2981 GTCGCCATATTCTACCCAGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 3040

Qy    2701 GGGCCTGCTTACCTCCTAAAAGGTGCTTTGACGCGCGTGCCGTACTTCGTCAGGGCTCAC 2760
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3041 GGGCCTGCTTACCTCCTAAAAGGTGCTTTGACGCGCGTGCCGTACTTCGTCAGGGCTCAC 3100

Qy    2761 GCTCTACTGAGGATGTGCACCATGGCAAGGCATCTCGCGGGGGGCAGGTACGTCCAGATG 2820
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3101 GCTCTACTGAGGATGTGCACCATGGCAAGGCATCTCGCGGGGGGCAGGTACGTCCAGATG 3160

Qy    2821 GCGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 2880
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3161 GCGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 3220

Qy    2881 TCGGATTGGGCTGCTAGTGGCCTGCGGGACCTGGCGGTCGCCGTTGAGCCTATCATCTTC 2940
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3221 TCGGATTGGGCTGCTAGTGGCCTGCGGGACCTGGCGGTCGCCGTTGAGCCTATCATCTTC 3280

Qy    2941 AGTCCGATGGAGAAGAAAGTCATTGTCTGGGGAGCGGAGACAGCTGCTTGTGGGGACATT 3000
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3281 AGTCCGATGGAGAAGAAAGTCATTGTCTGGGGAGCGGAGACAGCTGCTTGTGGGGACATT 3340

Qy    3001 TTACACGGACTTCCCGTGTCCGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3060
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3341 TTACACGGACTTCCCGTGTCCGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3400

Qy    3061 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3401 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3460

Qy    3121 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCT 3180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3461 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCT 3520

Qy    3181 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3521 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3580

Qy    3241 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3581 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3640

Qy    3301 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGG 3360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3641 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGG 3700

Qy    3361 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3701 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3760

Qy    3421 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3761 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3820

Qy    3481 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3540
```

```
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3821 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3880

Qy    3541 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3881 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3940

Qy    3601 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 3660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3941 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 4000

Qy    3661 CCTGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 3720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4001 CCTGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 4060

Qy    3721 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 3780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4061 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 4120

Qy    3781 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 3840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4121 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 4180

Qy    3841 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 3900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4181 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 4240

Qy    3901 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 3960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4241 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 4300

Qy    3961 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4301 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4360

Qy    4021 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4361 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4420

Qy    4081 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4421 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4480

Qy    4141 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4481 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4540

Qy    4201 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4541 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4600

Qy    4261 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4601 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4660

Qy    4321 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4661 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4720

Qy    4381 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4721 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4780

Qy    4441 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4781 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4840

Qy    4501 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4841 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4900

Qy    4561 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      4901 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4960
Qy      4621 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 4680
Db      4961 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 5020
Qy      4681 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 4740
Db      5021 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 5080
Qy      4741 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 4800
Db      5081 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 5140
Qy      4801 GCCAAAGCCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 4860
Db      5141 GCCAAAGCCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 5200
Qy      4861 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 4920
Db      5201 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 5260
Qy      4921 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 4980
Db      5261 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 5320
Qy      4981 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5040
Db      5321 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5380
Qy      5041 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5100
Db      5381 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5440
Qy      5101 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5160
Db      5441 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5500
Qy      5161 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5220
Db      5501 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5560
Qy      5221 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5280
Db      5561 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5620
Qy      5281 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5340
Db      5621 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5680
Qy      5341 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5400
Db      5681 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5740
Qy      5401 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5460
Db      5741 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5800
Qy      5461 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5520
Db      5801 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5860
Qy      5521 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5580
Db      5861 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5920
Qy      5581 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5640
Db      5921 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5980
Qy      5641 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 5700
Db      5981 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 6040

Qy      5701 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 5760
Db      6041 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 6100
Qy      5761 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 5820
Db      6101 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 6160
Qy      5821 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 5880
Db      6161 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 6220
Qy      5881 CTCAGAAGACTCCACAACTGGATTACTGAGGACTGCCCCATCCCATGCGGCGGCTCGTGG 5940
Db      6221 CTCAGAAGACTCCACAACTGGATTACTGAGGACTGCCCCATCCCATGCGGCGGCTCGTGG 6280
Qy      5941 CTCCGCGATGTGTGGGACTGGGTTTGCACCATCCTAACAGACTTTAAAAATTGGCTGACC 6000
Db      6281 CTCCGCGATGTGTGGGACTGGGTTTGCACCATCCTAACAGACTTTAAAAATTGGCTGACC 6340
Qy      6001 TCCAAATTATTCCCAAAGATGCCCGGCCTCCCCTTTGTCTCCTGTCAAAAGGGGTACAAG 6060
Db      6341 TCCAAATTATTCCCAAAGATGCCCGGCCTCCCCTTTGTCTCCTGTCAAAAGGGGTACAAG 6400
Qy      6061 GGCGTGTGGGCCGGCACTGGCATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6120
Db      6401 GGCGTGTGGGCCGGCACTGGCATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6460
Qy      6121 GGCAATGTCCGCTTGGGCTCCATGAGAATCACGGGGCCTAAGACCTGCATGAATATCTGG 6180
Db      6461 GGCAATGTCCGCTTGGGCTCCATGAGAATCACGGGGCCTAAGACCTGCATGAATATCTGG 6520
Qy      6181 CAGGGGACCTTTCCTATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCGCCA 6240
Db      6521 CAGGGGACCTTTCCTATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCGCCA 6580
Qy      6241 AACTTTAAGGTCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6300
Db      6581 AACTTTAAGGTCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6640
Qy      6301 GGGTCATACCACTACATAACAGGACTCACCACTGATAACTTGAAAGTCCCCTGCCAACTA 6360
Db      6641 GGGTCATACCACTACATAACAGGACTCACCACTGATAACTTGAAAGTCCCCTGCCAACTA 6700
Qy      6361 CCCTCTCCCGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCACA 6420
Db      6701 CCCTCTCCCGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCACA 6760
Qy      6421 CCGAAGCCGTTTTTCCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6480
Db      6761 CCGAAGCCGTTTTTCCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6820
Qy      6481 GGGTCCCAGCTTCCTTGCGACCCTGAACCCGACACAGACGTATTGATGTCCATGCTAACA 6540
Db      6821 GGGTCCCAGCTTCCTTGCGACCCTGAACCCGACACAGACGTATTGATGTCCATGCTAACA 6880
Qy      6541 GATCCATCTCATATCACGGCGGAGACTGCAGCGCGGCGTTTAGCGCGGGGGTCACCCCCA 6600
Db      6881 GATCCATCTCATATCACGGCGGAGACTGCAGCGCGGCGTTTAGCGCGGGGGTCACCCCCA 6940
Qy      6601 TCCGAGGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 6660
Db      6941 TCCGAGGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 7000
Qy      6661 ACCCACGGCAAAGCCTATGATGTGGACATGGTGGATGCTAACCTGTTCATGGGGGGCGAT 6720
Db      7001 ACCCACGGCAAAGCCTATGATGTGGACATGGTGGATGCTAACCTGTTCATGGGGGGCGAT 7060
Qy      6721 GTGACTCGGATAGAGTCTGGGTCCAAAGTGGTCGTTCTGGACTCTCTCGACCCAATGGTC 6780
Db      7061 GTGACTCGGATAGAGTCTGGGTCCAAAGTGGTCGTTCTGGACTCTCTCGACCCAATGGTC 7120
```

```
Qy      6781 GAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCAGAATACATGCTCCCCAAGAAGAGG 6840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7121 GAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCAGAATACATGCTCCCCAAGAAGAGG 7180

Qy      6841 TTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCGCTTGTGGAATCG 6900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7181 TTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCGCTTGTGGAATCG 7240

Qy      6901 TGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTCCCTCCTCCTAGG 6960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7241 TGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTCCCTCCTCCTAGG 7300

Qy      6961 AAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGTGAGGACTCCATA 7020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7301 AAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGTGAGGACTCCATA 7360

Qy      7021 GGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCCCCCAAGCGGCGAT 7080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7361 GGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCCCCCAAGCGGCGAT 7420

Qy      7081 TCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACGCCTCCTGATGAG 7140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7421 TCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACGCCTCCTGATGAG 7480

Qy      7141 TTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCTCGAGGGGGAGCTTGGA 7200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7481 TTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCTCGAGGGGGAGCTTGGA 7540

Qy      7201 GATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCCAGGGGGGGGTGGCA 7260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7541 GATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCCAGGGGGGGGTGGCA 7600

Qy      7261 GCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGACGACTCCGTCGTG 7320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7601 GCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGACGACTCCGTCGTG 7660

Qy      7321 TGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGTAGTCCCGAAGAG 7380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7661 TGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGTAGTCCCGAAGAG 7720

Qy      7381 GAGAAGTTACCGATTAACCCCTTGAGCAACTCCCTGTTGCGATATCACAACAAGGTGTAC 7440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7721 GAGAAGTTACCGATTAACCCCTTGAGCAACTCCCTGTTGCGATATCACAACAAGGTGTAC 7780

Qy      7441 TGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTTGATAGGATGCAA 7500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7781 TGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTTGATAGGATGCAA 7840

Qy      7501 GTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCGGCCTCCAAGGTC 7560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7841 GTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCGGCCTCCAAGGTC 7900

Qy      7561 ACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCCCATTCTGCAAGA 7620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7901 ACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCCCATTCTGCAAGA 7960

Qy      7621 TCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGGGCCGTTAACCAC 7680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7961 TCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGGGCCGTTAACCAC 8020

Qy      7681 ATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATTCCCACAACCATT 7740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8021 ATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATTCCCACAACCATT 8080

Qy      7741 ATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAGAAAGCAGCTCGC 7800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8081 ATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAGAAAGCAGCTCGC 8140

Qy      7801 CTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCCCTTTATGACATT 7860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8141 CTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCCCTTTATGACATT 8200

Qy      7861 ACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAGTATTCCCCCGCT 7920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8201 ACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAGTATTCCCCCGCT 8260

Qy      7921 CAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCTATGGGTTTTTCG 7980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8261 CAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCTATGGGTTTTTCG 8320

Qy      7981 TATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGGACTGAGGAGTCC 8040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8321 TATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGGACTGAGGAGTCC 8380

Qy      8041 ATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACACTCGCTAACTGAG 8100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8381 ATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACACTCGCTAACTGAG 8440

Qy      8101 AGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGCGGGTACAGGCGT 8160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8441 AGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGCGGGTACAGGCGT 8500

Qy      8161 TGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACATGCTACGTGAAA 8220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8501 TGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACATGCTACGTGAAA 8560

Qy      8221 GCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTGGTATGCGGCGAT 8280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8561 GCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTGGTATGCGGCGAT 8620

Qy      8281 GACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGGAACCTGAGAGCC 8340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8621 GACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGGAACCTGAGAGCC 8680

Qy      8341 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCCAGACCGGAGTAT 8400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8681 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCCAGACCGGAGTAT 8740

Qy      8401 GATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTGGGCCCACAAGGC 8460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8741 GATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTGGGCCCACAAGGC 8800

Qy      8461 CGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGGGCTGCCTGGGAA 8520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8801 CGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGGGCTGCCTGGGAA 8860

Qy      8521 ACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAGTACGCCCCGACC 8580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8861 ACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAGTACGCCCCGACC 8920

Qy      8581 ATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATGGCTCAAGACACG 8640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8921 ATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATGGCTCAAGACACG 8980

Qy      8641 CTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTGAGTCCCTTGGAC 8700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8981 CTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTGAGTCCCTTGGAC 9040

Qy      8701 CTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTGCACACATACACT 8760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      9041 CTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTGCACACATACACT 9100

Qy      8761 CCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGA 8820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      9101 CCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGA 9160

Qy      8821 GCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGTGGGGGGAGAGCG 8880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      9161 GCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGTGGGGGGAGAGCG 9220

Qy      8881 GCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 8940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      9221 GCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 9280

Qy      8941 TTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGCGCCGGCGGGGGC 9000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       9281 TTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGCGCCGGCGGGGGC 9340

Qy       9001 GACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTTGGCCTACTCCTA 9060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9341 GACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTTGGCCTACTCCTA 9400

Qy       9061 CTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCGG 9099
             |||||||||||||||||||||||||||||||||||||||
Db       9401 CTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCGG 9439


RESULT 2
AX057317
LOCUS       AX057317                9711 bp    DNA     linear   PAT 17-JAN-2001
DEFINITION  Sequence 3 from Patent WO0075352.
ACCESSION   AX057317
VERSION     AX057317.1  GI:12310062
KEYWORDS    .
SOURCE      Hepatitis C virus
  ORGANISM  Hepatitis C virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepacivirus.
REFERENCE   1
  AUTHORS   Nam,J.H., Bukh,J., Emerson,S.U. and Purcell,R.H.
  TITLE     Hcv/bvdv chimeric genomes and uses thereof
  JOURNAL   Patent: WO 0075352-A 3 14-DEC-2000;
            THE GOVERNMENT OF THE UNITED STATES OF AMERICA (US)
FEATURES             Location/Qualifiers
     source          1. .9711
                     /organism="Hepatitis C virus"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:11103"
ORIGIN

  Query Match             100.0%;  Score 9099;  DB 6;  Length 9711;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 9099;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 400

Qy         61 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 460

Qy        121 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 520

Qy        181 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 580

Qy        241 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 640

Qy        301 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        641 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 700

Qy        361 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        701 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 760

Qy        421 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        761 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 820

Qy        481 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        821 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 880

Qy        541 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        881 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 940

Qy        601 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        941 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 1000

Qy        661 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1001 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 1060

Qy        721 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1061 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 1120

Qy        781 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1121 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 1180

Qy        841 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1181 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 1240

Qy        901 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1241 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 1300

Qy        961 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1301 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1360

Qy       1021 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1361 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1420

Qy       1081 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1421 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1480

Qy       1141 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1481 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1540

Qy       1201 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1541 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1600

Qy       1261 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1601 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1660

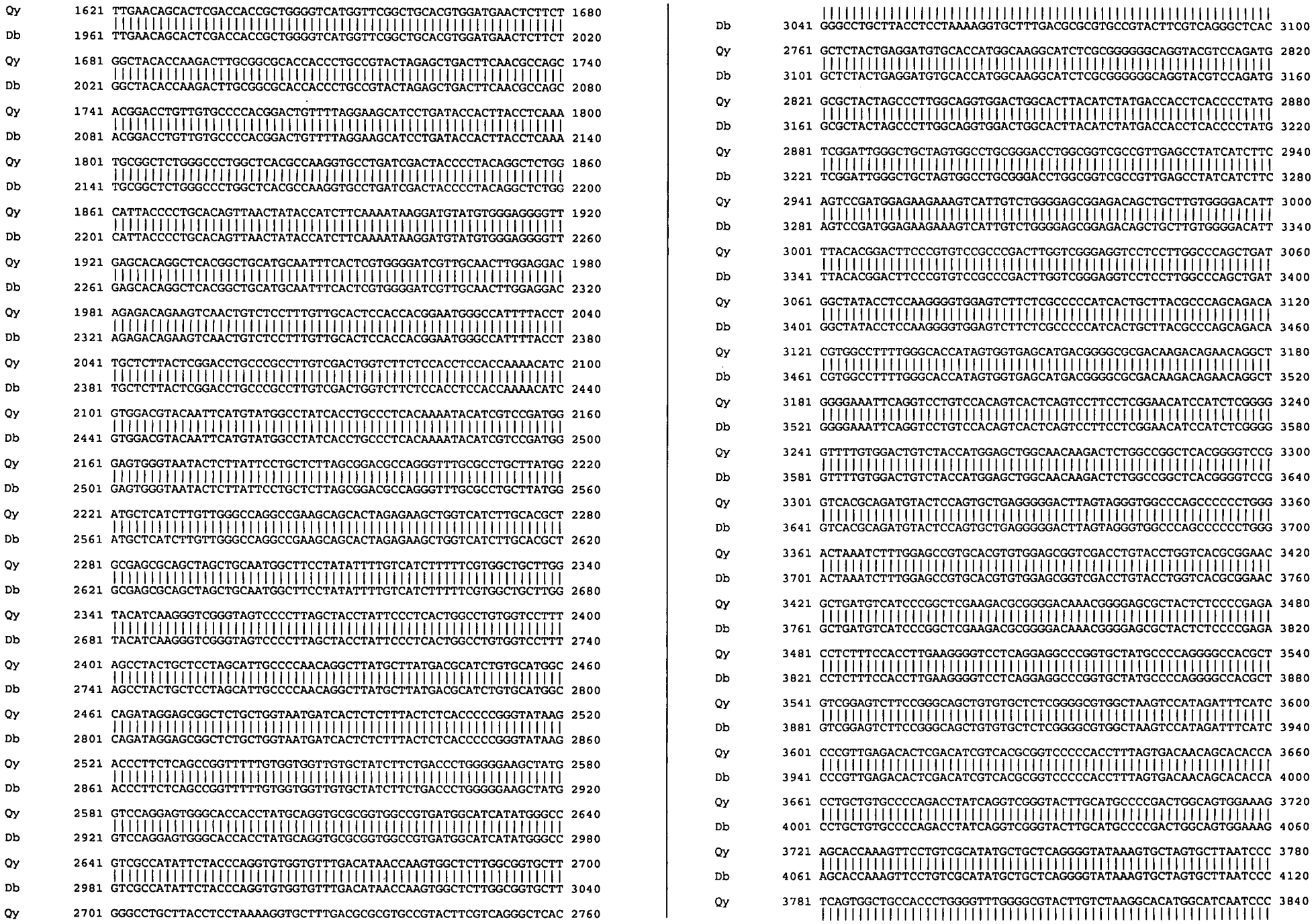
Qy       1321 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1661 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1720

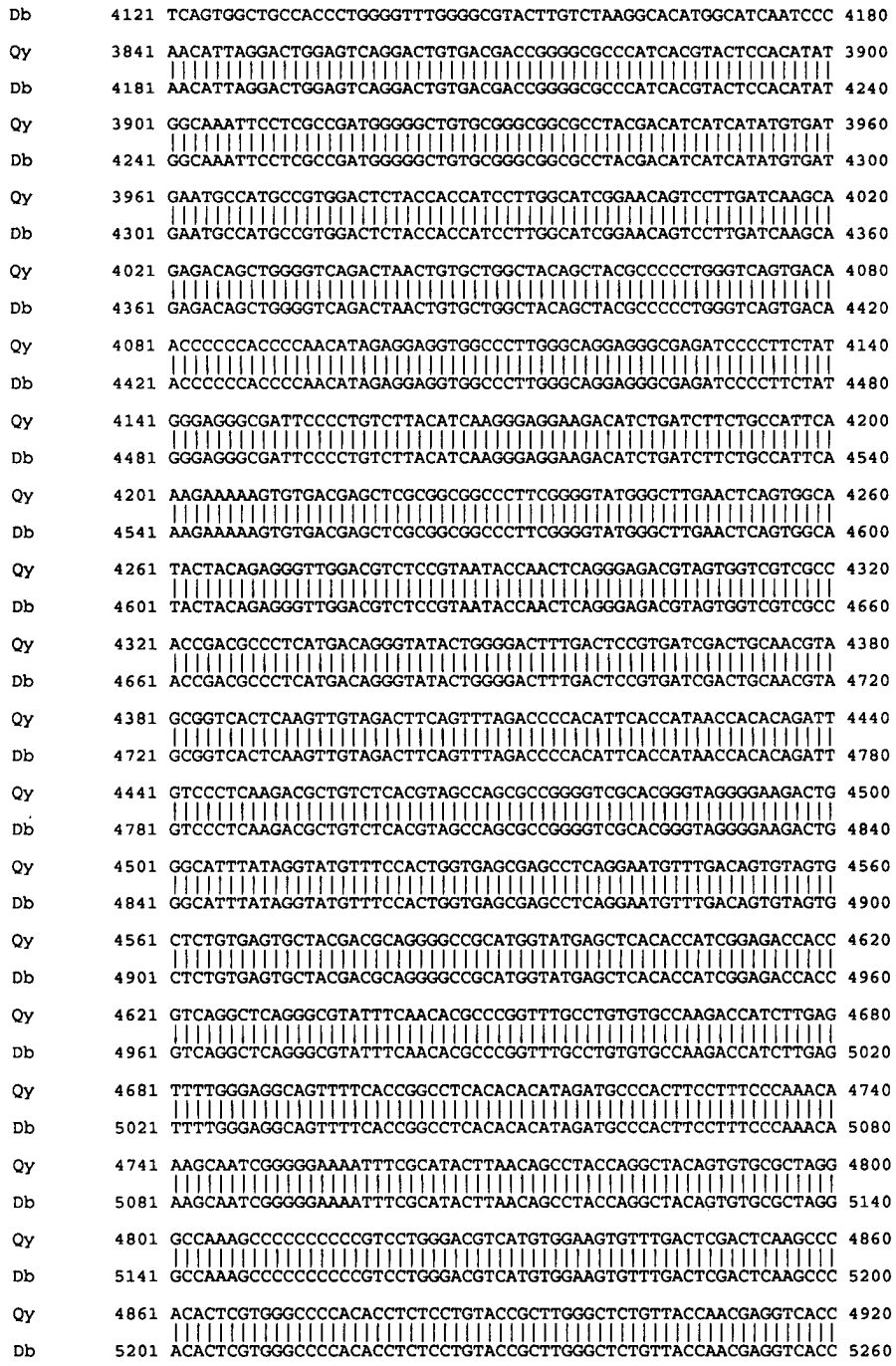
Qy       1381 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1721 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1780

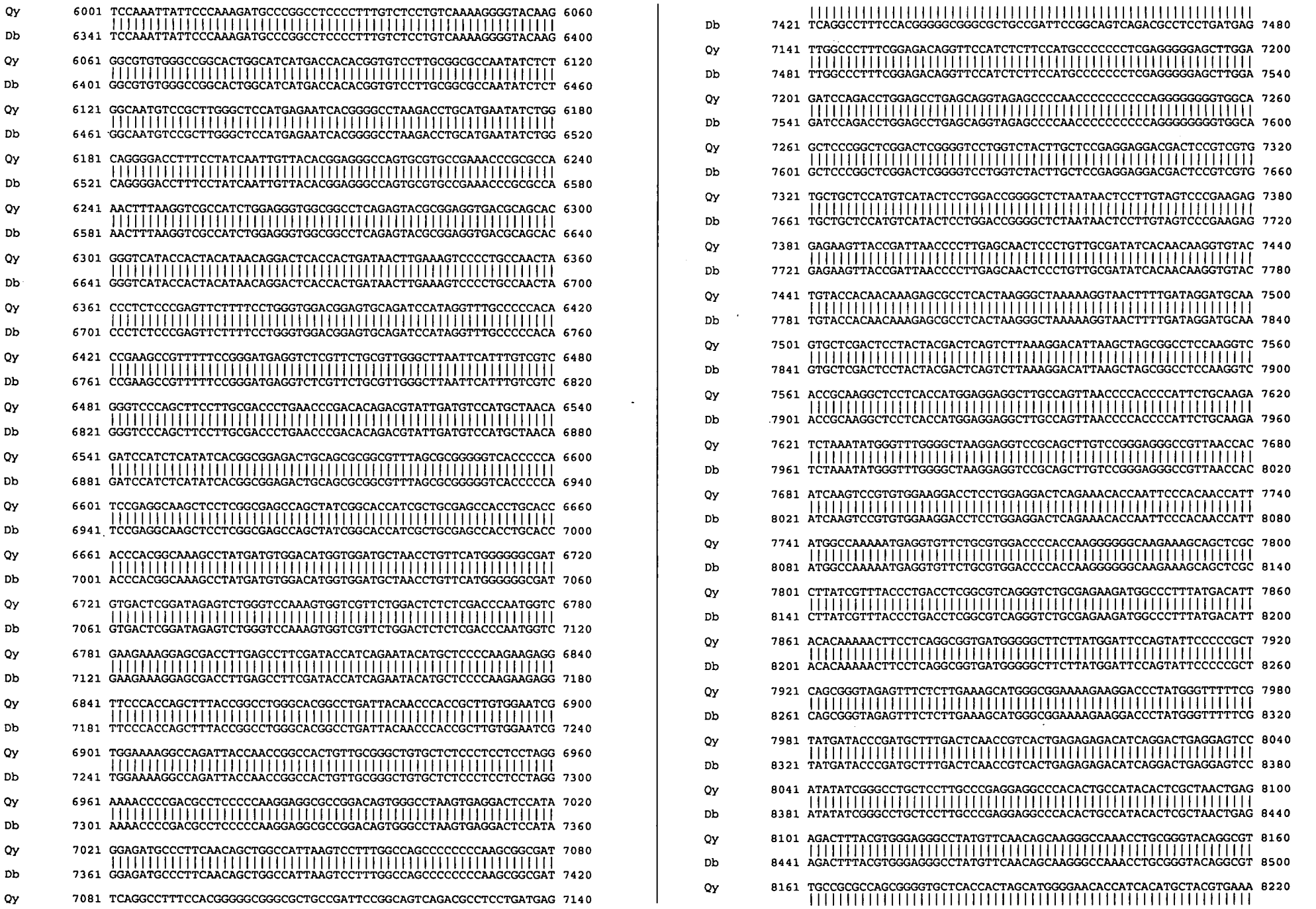
Qy       1441 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1781 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1840

Qy       1501 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1841 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1900

Qy       1561 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1901 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1960
```


Qy 1621 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 1680
Db 1961 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 2020
Qy 1681 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 1740
Db 2021 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 2080
Qy 1741 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 1800
Db 2081 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 2140
Qy 1801 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 1860
Db 2141 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 2200
Qy 1861 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 1920
Db 2201 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 2260
Qy 1921 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 1980
Db 2261 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 2320
Qy 1981 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2040
Db 2321 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2380
Qy 2041 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2100
Db 2381 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2440
Qy 2101 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2160
Db 2441 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2500
Qy 2161 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2220
Db 2501 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2560
Qy 2221 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2280
Db 2561 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2620
Qy 2281 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATATTTTGTCATCTTTTTCGTGGCTGCTTGG 2340
Db 2621 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATATTTTGTCATCTTTTTCGTGGCTGCTTGG 2680
Qy 2341 TACATCAAGGGTCGGGTAGTCCCCTTAGCTACCTATTCCCTCACTGGCCTGTGGTCCTTT 2400
Db 2681 TACATCAAGGGTCGGGTAGTCCCCTTAGCTACCTATTCCCTCACTGGCCTGTGGTCCTTT 2740
Qy 2401 AGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTGCATGGC 2460
Db 2741 AGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTGCATGGC 2800
Qy 2461 CAGATAGGAGCGGCTCTGCTGGTAATGATCACTCTCTTTACTCTCACCCCCGGGTATAAG 2520
Db 2801 CAGATAGGAGCGGCTCTGCTGGTAATGATCACTCTCTTTACTCTCACCCCCGGGTATAAG 2860
Qy 2521 ACCCTTCTCAGCCGGTTTTTGTGGTGGTTGTGCTATCTTCTGACCCTGGGGGAAGCTATG 2580
Db 2861 ACCCTTCTCAGCCGGTTTTTGTGGTGGTTGTGCTATCTTCTGACCCTGGGGGAAGCTATG 2920
Qy 2581 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGCATCATATGGGCC 2640
Db 2921 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGCATCATATGGGCC 2980
Qy 2641 GTCGCCATATTCTACCCAGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 2700
Db 2981 GTCGCCATATTCTACCCAGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 3040
Qy 2701 GGGCCTGCTTACCTCCTAAAAGGTGCTTTGACGCGCGTGCCGTACTTCGTCAGGGCTCAC 2760
Db 3041 GGGCCTGCTTACCTCCTAAAAGGTGCTTTGACGCGCGTGCCGTACTTCGTCAGGGCTCAC 3100
Qy 2761 GCTCTACTGAGGATGTGCACCATGGCAAGGCATCTCGCGGGGGGCAGGTACGTCCAGATG 2820
Db 3101 GCTCTACTGAGGATGTGCACCATGGCAAGGCATCTCGCGGGGGGCAGGTACGTCCAGATG 3160
Qy 2821 GCGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 2880
Db 3161 GCGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 3220
Qy 2881 TCGGATTGGGCTGCTAGTGGCCTGCGGGACCTGGCGGTCGCCGTTGAGCCTATCATCTTC 2940
Db 3221 TCGGATTGGGCTGCTAGTGGCCTGCGGGACCTGGCGGTCGCCGTTGAGCCTATCATCTTC 3280
Qy 2941 AGTCCGATGGAGAAGAAAGTCATTGTCTGGGGAGCGGAGACAGCTGCTTGTGGGGACATT 3000
Db 3281 AGTCCGATGGAGAAGAAAGTCATTGTCTGGGGAGCGGAGACAGCTGCTTGTGGGGACATT 3340
Qy 3001 TTACACGGACTTCCCGTGTCCGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3060
Db 3341 TTACACGGACTTCCCGTGTCCGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3400
Qy 3061 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3120
Db 3401 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3460
Qy 3121 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCT 3180
Db 3461 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCT 3520
Qy 3181 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3240
Db 3521 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3580
Qy 3241 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3300
Db 3581 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3640
Qy 3301 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGG 3360
Db 3641 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGG 3700
Qy 3361 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3420
Db 3701 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3760
Qy 3421 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3480
Db 3761 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3820
Qy 3481 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3540
Db 3821 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3880
Qy 3541 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3600
Db 3881 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3940
Qy 3601 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 3660
Db 3941 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 4000
Qy 3661 CCTGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 3720
Db 4001 CCTGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 4060
Qy 3721 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 3780
Db 4061 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 4120
Qy 3781 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 3840


Db 4121 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 4180
Qy 3841 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 3900
Db 4181 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 4240
Qy 3901 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 3960
Db 4241 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 4300
Qy 3961 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4020
Db 4301 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4360
Qy 4021 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4080
Db 4361 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4420
Qy 4081 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4140
Db 4421 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4480
Qy 4141 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4200
Db 4481 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4540
Qy 4201 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4260
Db 4541 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4600
Qy 4261 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4320
Db 4601 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4660
Qy 4321 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4380
Db 4661 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4720
Qy 4381 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4440
Db 4721 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4780
Qy 4441 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4500
Db 4781 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4840
Qy 4501 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4560
Db 4841 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4900
Qy 4561 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4620
Db 4901 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4960
Qy 4621 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 4680
Db 4961 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 5020
Qy 4681 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 4740
Db 5021 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 5080
Qy 4741 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 4800
Db 5081 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 5140
Qy 4801 GCCAAAGCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 4860
Db 5141 GCCAAAGCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 5200
Qy 4861 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 4920
Db 5201 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 5260
Qy 4921 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 4980
Db 5261 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 5320
Qy 4981 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5040
Db 5321 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5380
Qy 5041 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5100
Db 5381 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5440
Qy 5101 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5160
Db 5441 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5500
Qy 5161 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5220
Db 5501 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5560
Qy 5221 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5280
Db 5561 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5620
Qy 5281 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5340
Db 5621 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5680
Qy 5341 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5400
Db 5681 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5740
Qy 5401 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5460
Db 5741 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5800
Qy 5461 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5520
Db 5801 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5860
Qy 5521 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5580
Db 5861 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5920
Qy 5581 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5640
Db 5921 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5980
Qy 5641 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 5700
Db 5981 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 6040
Qy 5701 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 5760
Db 6041 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 6100
Qy 5761 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 5820
Db 6101 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 6160
Qy 5821 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 5880
Db 6161 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 6220
Qy 5881 CTCAGAAGACTCCACAACTGGATTACTGAGGACTGCCCCATCCCATGCGGCGGCTCGTGG 5940
Db 6221 CTCAGAAGACTCCACAACTGGATTACTGAGGACTGCCCCATCCCATGCGGCGGCTCGTGG 6280
Qy 5941 CTCCGCGATGTGTGGGACTGGGTTTGCACCATCCTAACAGACTTTAAAAATTGGCTGACC 6000
Db 6281 CTCCGCGATGTGTGGGACTGGGTTTGCACCATCCTAACAGACTTTAAAAATTGGCTGACC 6340


Qy 6001 TCCAAATTATTCCCAAAGATGCCCGGCCTCCCCTTTGTCTCCTGTCAAAAGGGGTACAAG 6060
Db 6341 TCCAAATTATTCCCAAAGATGCCCGGCCTCCCCTTTGTCTCCTGTCAAAAGGGGTACAAG 6400
Qy 6061 GGCGTGTGGGCCGGCACTGGCATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6120
Db 6401 GGCGTGTGGGCCGGCACTGGCATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6460
Qy 6121 GGCAATGTCCGCTTGGGCTCCATGAGAATCACGGGGCCTAAGACCTGCATGAATATCTGG 6180
Db 6461 GGCAATGTCCGCTTGGGCTCCATGAGAATCACGGGGCCTAAGACCTGCATGAATATCTGG 6520
Qy 6181 CAGGGGACCTTTCCTATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCGCCA 6240
Db 6521 CAGGGGACCTTTCCTATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCGCCA 6580
Qy 6241 AACTTTAAGGTCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6300
Db 6581 AACTTTAAGGTCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6640
Qy 6301 GGGTCATACCACTACATAACAGGACTCACCACTGATAACTTGAAAGTCCCCTGCCAACTA 6360
Db 6641 GGGTCATACCACTACATAACAGGACTCACCACTGATAACTTGAAAGTCCCCTGCCAACTA 6700
Qy 6361 CCCTCTCCCGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCACA 6420
Db 6701 CCCTCTCCCGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCACA 6760
Qy 6421 CCGAAGCCGTTTTTCCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6480
Db 6761 CCGAAGCCGTTTTTCCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6820
Qy 6481 GGGTCCCAGCTTCCTTGCGACCCTGAACCCGACACAGACGTATTGATGTCCATGCTAACA 6540
Db 6821 GGGTCCCAGCTTCCTTGCGACCCTGAACCCGACACAGACGTATTGATGTCCATGCTAACA 6880
Qy 6541 GATCCATCTCATATCACGGCGGAGACTGCAGCGCGGCGTTTAGCGCGGGGGTCACCCCCA 6600
Db 6881 GATCCATCTCATATCACGGCGGAGACTGCAGCGCGGCGTTTAGCGCGGGGGTCACCCCCA 6940
Qy 6601 TCCGAGGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 6660
Db 6941 TCCGAGGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 7000
Qy 6661 ACCCACGGCAAAGCCTATGATGTGGACATGGTGGATGCTAACCTGTTCATGGGGGGCGAT 6720
Db 7001 ACCCACGGCAAAGCCTATGATGTGGACATGGTGGATGCTAACCTGTTCATGGGGGGCGAT 7060
Qy 6721 GTGACTCGGATAGAGTCTGGGTCCAAAGTGGTCGTTCTGGACTCTCTCGACCCAATGGTC 6780
Db 7061 GTGACTCGGATAGAGTCTGGGTCCAAAGTGGTCGTTCTGGACTCTCTCGACCCAATGGTC 7120
Qy 6781 GAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCAGAATACATGCTCCCCAAGAAGAGG 6840
Db 7121 GAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCAGAATACATGCTCCCCAAGAAGAGG 7180
Qy 6841 TTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCGCTTGTGGAATCG 6900
Db 7181 TTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCGCTTGTGGAATCG 7240
Qy 6901 TGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTCCCTCCTCCTAGG 6960
Db 7241 TGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTCCCTCCTCCTAGG 7300
Qy 6961 AAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGTGAGGACTCCATA 7020
Db 7301 AAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGTGAGGACTCCATA 7360
Qy 7021 GGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCCCCCAAGCGGCGAT 7080
Db 7361 GGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCCCCAAGCGGCGAT 7420
Qy 7081 TCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACGCCTCCTGATGAG 7140
Db 7421 TCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACGCCTCCTGATGAG 7480
Qy 7141 TTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCTCGAGGGGGAGCTTGGA 7200
Db 7481 TTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCTCGAGGGGGAGCTTGGA 7540
Qy 7201 GATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCCAGGGGGGGGTGGCA 7260
Db 7541 GATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCAGGGGGGGGTGGCA 7600
Qy 7261 GCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGACGACTCCGTCGTG 7320
Db 7601 GCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGACGACTCCGTCGTG 7660
Qy 7321 TGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGTAGTCCCGAAGAG 7380
Db 7661 TGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGTAGTCCCGAAGAG 7720
Qy 7381 GAGAAGTTACCGATTAACCCCTTGAGCAACTCCCTGTTGCGATATCACAACAAGGTGTAC 7440
Db 7721 GAGAAGTTACCGATTAACCCCTTGAGCAACTCCCTGTTGCGATATCACAACAAGGTGTAC 7780
Qy 7441 TGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTTGATAGGATGCAA 7500
Db 7781 TGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTTGATAGGATGCAA 7840
Qy 7501 GTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCGGCCTCCAAGGTC 7560
Db 7841 GTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCGGCCTCCAAGGTC 7900
Qy 7561 ACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCCCATTCTGCAAGA 7620
Db 7901 ACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCCCATTCTGCAAGA 7960
Qy 7621 TCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGGGCCGTTAACCAC 7680
Db 7961 TCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGGGCCGTTAACCAC 8020
Qy 7681 ATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATTCCCACAACCATT 7740
Db 8021 ATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATTCCCACAACCATT 8080
Qy 7741 ATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAGAAAGCAGCTCGC 7800
Db 8081 ATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAGAAAGCAGCTCGC 8140
Qy 7801 CTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCCCTTTATGACATT 7860
Db 8141 CTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCCCTTTATGACATT 8200
Qy 7861 ACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAGTATTCCCCCGCT 7920
Db 8201 ACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAGTATTCCCCCGCT 8260
Qy 7921 CAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCTATGGGTTTTTCG 7980
Db 8261 CAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCTATGGGTTTTTCG 8320
Qy 7981 TATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGGACTGAGGAGTCC 8040
Db 8321 TATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGGACTGAGGAGTCC 8380
Qy 8041 ATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACACTCGCTAACTGAG 8100
Db 8381 ATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACACTCGCTAACTGAG 8440
Qy 8101 AGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGCGGGTACAGGCGT 8160
Db 8441 AGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGCGGGTACAGGCGT 8500
Qy 8161 TGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACATGCTACGTGAAA 8220

```
Db       8501 TGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACATGCTACGTGAAA 8560

Qy       8221 GCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTGGTATGCGGCGAT 8280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8561 GCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTGGTATGCGGCGAT 8620

Qy       8281 GACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGGAACCTGAGAGCC 8340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8621 GACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGGAACCTGAGAGCC 8680

Qy       8341 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCCCAGACCGGAGTAT 8400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8681 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCCCAGACCGGAGTAT 8740

Qy       8401 GATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTGGGCCCACAAGGC 8460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8741 GATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTGGGCCCACAAGGC 8800

Qy       8461 CGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGGGCTGCCTGGGAA 8520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8801 CGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGGGCTGCCTGGGAA 8860

Qy       8521 ACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAGTACGCCCCGACC 8580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8861 ACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAGTACGCCCCGACC 8920

Qy       8581 ATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATGGCTCAAGACACG 8640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8921 ATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATGGCTCAAGACACG 8980

Qy       8641 CTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTGAGTCCCTTGGAC 8700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8981 CTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTGAGTCCCTTGGAC 9040

Qy       8701 CTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTGCACACATACACT 8760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9041 CTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTGCACACATACACT 9100

Qy       8761 CCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGA 8820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9101 CCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGA 9160

Qy       8821 GCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGTGGGGGGAGAGCG 8880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9161 GCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGTGGGGGGAGAGCG 9220

Qy       8881 GCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 8940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9221 GCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 9280

Qy       8941 TTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGCGCCGGCGGGGGC 9000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9281 TTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGCGCCGGCGGGGGC 9340

Qy       9001 GACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTTGGCCTACTCCTA 9060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9341 GACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTTGGCCTACTCCTA 9400

Qy       9061 CTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCGG 9099
             |||||||||||||||||||||||||||||||||||||||
Db       9401 CTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCGG 9439


RESULT 3
AX057395
LOCUS       AX057395               9711 bp    DNA     linear   PAT 17-JAN-2001
DEFINITION  Sequence 4 from Patent WO0075337.
ACCESSION   AX057395
VERSION     AX057395.1  GI:12310134
KEYWORDS    .
SOURCE      Hepatitis C virus
  ORGANISM  Hepatitis C virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepacivirus.
REFERENCE   1
  AUTHORS   Bukh,J., Yanagi,M., Emerson,S.U. and Purcell,R.H.
  TITLE     Infectious cdna clone of gb virus b and uses thereof
  JOURNAL   Patent: WO 0075337-A 4 14-DEC-2000;
            THE GOVERNMENT OF THE UNITED STATES OF AMERICA (US)
FEATURES             Location/Qualifiers
     source          1. .9711
                     /organism="Hepatitis C virus"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:11103"
ORIGIN

  Query Match             100.0%;  Score 9099;  DB 6;  Length 9711;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 9099;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 400

Qy         61 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 460

Qy        121 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 520

Qy        181 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 580

Qy        241 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 640

Qy        301 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        641 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 700

Qy        361 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        701 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 760

Qy        421 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        761 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 820

Qy        481 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        821 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 880

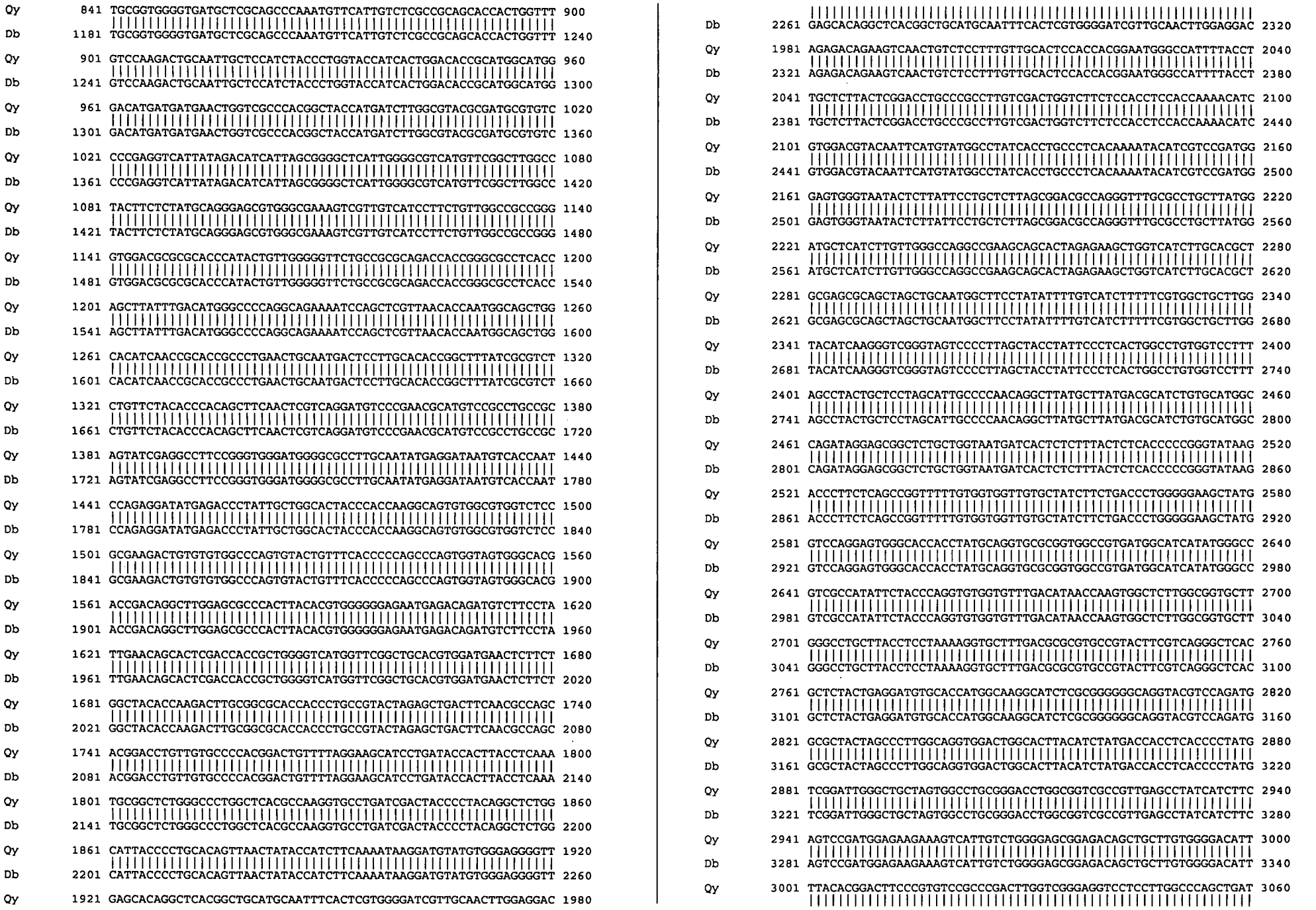
Qy        541 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        881 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 940

Qy        601 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        941 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 1000

Qy        661 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1001 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 1060

Qy        721 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1061 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 1120

Qy        781 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1121 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 1180
```


Qy 841 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 900
Db 1181 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 1240
Qy 901 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 960
Db 1241 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 1300
Qy 961 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1020
Db 1301 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1360
Qy 1021 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1080
Db 1361 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1420
Qy 1081 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1140
Db 1421 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1480
Qy 1141 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1200
Db 1481 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1540
Qy 1201 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1260
Db 1541 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1600
Qy 1261 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1320
Db 1601 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1660
Qy 1321 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1380
Db 1661 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1720
Qy 1381 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1440
Db 1721 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1780
Qy 1441 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1500
Db 1781 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1840
Qy 1501 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1560
Db 1841 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1900
Qy 1561 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1620
Db 1901 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1960
Qy 1621 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 1680
Db 1961 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 2020
Qy 1681 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 1740
Db 2021 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 2080
Qy 1741 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 1800
Db 2081 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 2140
Qy 1801 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 1860
Db 2141 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 2200
Qy 1861 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 1920
Db 2201 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 2260
Qy 1921 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 1980
Db 2261 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 2320
Qy 1981 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2040
Db 2321 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2380
Qy 2041 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2100
Db 2381 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2440
Qy 2101 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2160
Db 2441 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2500
Qy 2161 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2220
Db 2501 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2560
Qy 2221 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2280
Db 2561 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2620
Qy 2281 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATATTTTGTCATCTTTTTCGTGGCTGCTTGG 2340
Db 2621 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATATTTTGTCATCTTTTTCGTGGCTGCTTGG 2680
Qy 2341 TACATCAAGGGTCGGGTAGTCCCCTTAGCTACCTATTCCCTCACTGGCCTGTGGTCCTTT 2400
Db 2681 TACATCAAGGGTCGGGTAGTCCCCTTAGCTACCTATTCCCTCACTGGCCTGTGGTCCTTT 2740
Qy 2401 AGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTGCATGGC 2460
Db 2741 AGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTGCATGGC 2800
Qy 2461 CAGATAGGAGCGGCTCTGCTGGTAATGATCACTCTCTTTACTCTCACCCCCGGGTATAAG 2520
Db 2801 CAGATAGGAGCGGCTCTGCTGGTAATGATCACTCTCTTTACTCTCACCCCCGGGTATAAG 2860
Qy 2521 ACCCTTCTCAGCCGGTTTTTGTGGTGGTTGTGCTATCTTCTGACCCTGGGGGAAGCTATG 2580
Db 2861 ACCCTTCTCAGCCGGTTTTTGTGGTGGTTGTGCTATCTTCTGACCCTGGGGGAAGCTATG 2920
Qy 2581 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGCATCATATGGGCC 2640
Db 2921 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGCATCATATGGGCC 2980
Qy 2641 GTCGCCATATTCTACCCAGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 2700
Db 2981 GTCGCCATATTCTACCCAGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 3040
Qy 2701 GGGCCTGCTTACCTCCTAAAAGGTGCTTTGACGCGCGTGCCGTACTTCGTCAGGGCTCAC 2760
Db 3041 GGGCCTGCTTACCTCCTAAAAGGTGCTTTGACGCGCGTGCCGTACTTCGTCAGGGCTCAC 3100
Qy 2761 GCTCTACTGAGGATGTGCACCATGGCAAGGCATCTCGCGGGGGGCAGGTACGTCCAGATG 2820
Db 3101 GCTCTACTGAGGATGTGCACCATGGCAAGGCATCTCGCGGGGGGCAGGTACGTCCAGATG 3160
Qy 2821 GCGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 2880
Db 3161 GCGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 3220
Qy 2881 TCGGATTGGGCTGCTAGTGGCCTGCGGGACCTGGCGGTCGCCGTTGAGCCTATCATCTTC 2940
Db 3221 TCGGATTGGGCTGCTAGTGGCCTGCGGGACCTGGCGGTCGCCGTTGAGCCTATCATCTTC 3280
Qy 2941 AGTCCGATGGAGAAGAAAGTCATTGTCTGGGGAGCGGAGACAGCTGCTTGTGGGGACATT 3000
Db 3281 AGTCCGATGGAGAAGAAAGTCATTGTCTGGGGAGCGGAGACAGCTGCTTGTGGGGACATT 3340
Qy 3001 TTACACGGACTTCCCGTGTCCGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3060

```
Db      3341 TTACACGGACTTCCCGTGTCCGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3400
Qy      3061 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3120
Db      3401 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3460
Qy      3121 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCT 3180
Db      3461 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCT 3520
Qy      3181 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3240
Db      3521 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3580
Qy      3241 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3300
Db      3581 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3640
Qy      3301 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGG 3360
Db      3641 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGG 3700
Qy      3361 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3420
Db      3701 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3760
Qy      3421 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3480
Db      3761 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3820
Qy      3481 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3540
Db      3821 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3880
Qy      3541 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3600
Db      3881 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3940
Qy      3601 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 3660
Db      3941 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 4000
Qy      3661 CCTGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 3720
Db      4001 CCTGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 4060
Qy      3721 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 3780
Db      4061 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 4120
Qy      3781 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 3840
Db      4121 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 4180
Qy      3841 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 3900
Db      4181 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 4240
Qy      3901 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 3960
Db      4241 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 4300
Qy      3961 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4020
Db      4301 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4360
Qy      4021 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4080
Db      4361 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4420
Qy      4081 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4140
Db      4421 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4480

Qy      4141 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4200
Db      4481 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4540
Qy      4201 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4260
Db      4541 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4600
Qy      4261 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4320
Db      4601 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4660
Qy      4321 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4380
Db      4661 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4720
Qy      4381 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4440
Db      4721 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4780
Qy      4441 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4500
Db      4781 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4840
Qy      4501 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4560
Db      4841 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4900
Qy      4561 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4620
Db      4901 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4960
Qy      4621 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 4680
Db      4961 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 5020
Qy      4681 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 4740
Db      5021 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 5080
Qy      4741 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 4800
Db      5081 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 5140
Qy      4801 GCCAAAGCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 4860
Db      5141 GCCAAAGCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 5200
Qy      4861 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 4920
Db      5201 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 5260
Qy      4921 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 4980
Db      5261 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 5320
Qy      4981 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5040
Db      5321 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5380
Qy      5041 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5100
Db      5381 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5440
Qy      5101 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5160
Db      5441 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5500
Qy      5161 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5220
Db      5501 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5560
```

```
Qy      5221 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5561 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5620

Qy      5281 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5621 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5680

Qy      5341 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5681 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5740

Qy      5401 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5741 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5800

Qy      5461 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5801 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5860

Qy      5521 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5861 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5920

Qy      5581 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5921 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5980

Qy      5641 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 5700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5981 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 6040

Qy      5701 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 5760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6041 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 6100

Qy      5761 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 5820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6101 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 6160

Qy      5821 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 5880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6161 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 6220

Qy      5881 CTCAGAAGACTCCACAACTGGATTACTGAGGACTGCCCCATCCCATGCGGCGGCTCGTGG 5940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6221 CTCAGAAGACTCCACAACTGGATTACTGAGGACTGCCCCATCCCATGCGGCGGCTCGTGG 6280

Qy      5941 CTCCGCGATGTGTGGGACTGGGTTTGCACCATCCTAACAGACTTTAAAAATTGGCTGACC 6000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6281 CTCCGCGATGTGTGGGACTGGGTTTGCACCATCCTAACAGACTTTAAAAATTGGCTGACC 6340

Qy      6001 TCCAAATTATTCCCAAAGATGCCCGGCCTCCCCTTTGTCTCCTGTCAAAAGGGGTACAAG 6060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6341 TCCAAATTATTCCCAAAGATGCCCGGCCTCCCCTTTGTCTCCTGTCAAAAGGGGTACAAG 6400

Qy      6061 GGCGTGTGGGCCGGCACTGGCATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6401 GGCGTGTGGGCCGGCACTGGCATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6460

Qy      6121 GGCAATGTCCGCTTGGGCTCCATGAGAATCACGGGGCCTAAGACCTGCATGAATATCTGG 6180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6461 GGCAATGTCCGCTTGGGCTCCATGAGAATCACGGGGCCTAAGACCTGCATGAATATCTGG 6520

Qy      6181 CAGGGGACCTTTCCTATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCGCCA 6240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6521 CAGGGGACCTTTCCTATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCGCCA 6580

Qy      6241 AACTTTAAGGTCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6581 AACTTTAAGGTCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6640

Qy      6301 GGGTCATACCACTACATAACAGGACTCACCACTGATAACTTGAAAGTCCCCTGCCAACTA 6360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6641 GGGTCATACCACTACATAACAGGACTCACCACTGATAACTTGAAAGTCCCCTGCCAACTA 6700

Qy      6361 CCCTCTCCCGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCACA 6420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6701 CCCTCTCCCGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCACA 6760

Qy      6421 CCGAAGCCGTTTTTCCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6761 CCGAAGCCGTTTTTCCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6820

Qy      6481 GGGTCCCAGCTTCCTTGCGACCCTGAACCCGACACAGACGTATTGATGTCCATGCTAACA 6540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6821 GGGTCCCAGCTTCCTTGCGACCCTGAACCCGACACAGACGTATTGATGTCCATGCTAACA 6880

Qy      6541 GATCCATCTCATATCACGGCGGAGACTGCAGCGCGGCGTTTAGCGCGGGGGTCACCCCCA 6600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6881 GATCCATCTCATATCACGGCGGAGACTGCAGCGCGGCGTTTAGCGCGGGGGTCACCCCCA 6940

Qy      6601 TCCGAGGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 6660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6941 TCCGAGGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 7000

Qy      6661 ACCCACGGCAAAGCCTATGATGTGGACATGGTGGATGCTAACCTGTTCATGGGGGGCGAT 6720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7001 ACCCACGGCAAAGCCTATGATGTGGACATGGTGGATGCTAACCTGTTCATGGGGGGCGAT 7060

Qy      6721 GTGACTCGGATAGAGTCTGGGTCCAAAGTGGTCGTTCTGGACTCTCTCGACCCAATGGTC 6780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7061 GTGACTCGGATAGAGTCTGGGTCCAAAGTGGTCGTTCTGGACTCTCTCGACCCAATGGTC 7120

Qy      6781 GAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCAGAATACATGCTCCCCAAGAAGAGG 6840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7121 GAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCAGAATACATGCTCCCCAAGAAGAGG 7180

Qy      6841 TTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCGCTTGTGGAATCG 6900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7181 TTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCGCTTGTGGAATCG 7240

Qy      6901 TGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTCCCTCCTCCTAGG 6960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7241 TGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTCCCTCCTCCTAGG 7300

Qy      6961 AAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGTGAGGACTCCATA 7020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7301 AAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGTGAGGACTCCATA 7360

Qy      7021 GGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCCCCCAAGCGGCGAT 7080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7361 GGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCCCCCAAGCGGCGAT 7420

Qy      7081 TCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACGCCTCCTGATGAG 7140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7421 TCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACGCCTCCTGATGAG 7480

Qy      7141 TTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCTCGAGGGGGAGCTTGGA 7200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7481 TTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCTCGAGGGGGAGCTTGGA 7540

Qy      7201 GATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCCAGGGGGGGGTGGCA 7260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7541 GATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCCAGGGGGGGGTGGCA 7600

Qy      7261 GCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGACGACTCCGTCGTG 7320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7601 GCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGACGACTCCGTCGTG 7660

Qy      7321 TGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGTAGTCCCGAAGAG 7380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7661 TGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGTAGTCCCGAAGAG 7720

Qy      7381 GAGAAGTTACCGATTAACCCCTTGAGCAACTCCCTGTTGCGATATCACAACAAGGTGTAC 7440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       7721 GAGAAGTTACCGATTAACCCCTTGAGCAACTCCCTGTTGCGATATCACAACAAGGTGTAC 7780
Qy       7441 TGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTTGATAGGATGCAA 7500
Db       7781 TGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTTGATAGGATGCAA 7840

Qy       7501 GTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCGGCCTCCAAGGTC 7560
Db       7841 GTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCGGCCTCCAAGGTC 7900

Qy       7561 ACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCCCATTCTGCAAGA 7620
Db       7901 ACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCCCATTCTGCAAGA 7960

Qy       7621 TCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGGGCCGTTAACCAC 7680
Db       7961 TCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGGGCCGTTAACCAC 8020

Qy       7681 ATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATTCCCACAACCATT 7740
Db       8021 ATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATTCCCACAACCATT 8080

Qy       7741 ATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAGAAAGCAGCTCGC 7800
Db       8081 ATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAGAAAGCAGCTCGC 8140

Qy       7801 CTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCCCTTTATGACATT 7860
Db       8141 CTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCCCTTTATGACATT 8200

Qy       7861 ACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAGTATTCCCCCGCT 7920
Db       8201 ACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAGTATTCCCCCGCT 8260

Qy       7921 CAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCTATGGGTTTTTCG 7980
Db       8261 CAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCTATGGGTTTTTCG 8320

Qy       7981 TATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGGACTGAGGAGTCC 8040
Db       8321 TATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGGACTGAGGAGTCC 8380

Qy       8041 ATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACACTCGCTAACTGAG 8100
Db       8381 ATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACACTCGCTAACTGAG 8440

Qy       8101 AGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGCGGGTACAGGCGT 8160
Db       8441 AGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGCGGGTACAGGCGT 8500

Qy       8161 TGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACATGCTACGTGAAA 8220
Db       8501 TGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACATGCTACGTGAAA 8560

Qy       8221 GCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTGGTATGCGGCGAT 8280
Db       8561 GCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTGGTATGCGGCGAT 8620

Qy       8281 GACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGGAACCTGAGAGCC 8340
Db       8621 GACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGGAACCTGAGAGCC 8680

Qy       8341 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCCAGACCGGAGTAT 8400
Db       8681 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCCAGACCGGAGTAT 8740

Qy       8401 GATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTGGGCCCACAAGGC 8460
Db       8741 GATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTGGGCCCACAAGGC 8800

Qy       8461 CGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGGGCTGCCTGGGAA 8520
Db       8801 CGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGGGCTGCCTGGGAA 8860

Qy       8521 ACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAGTACGCCCCGACC 8580
Db       8861 ACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAGTACGCCCCGACC 8920

Qy       8581 ATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATGGCTCAAGACACG 8640
Db       8921 ATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATGGCTCAAGACACG 8980

Qy       8641 CTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTGAGTCCCTTGGAC 8700
Db       8981 CTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTGAGTCCCTTGGAC 9040

Qy       8701 CTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTGCACACATACACT 8760
Db       9041 CTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTGCACACATACACT 9100

Qy       8761 CCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGA 8820
Db       9101 CCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGA 9160

Qy       8821 GCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGTGGGGGGAGAGCG 8880
Db       9161 GCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGTGGGGGGAGAGCG 9220

Qy       8881 GCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 8940
Db       9221 GCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 9280

Qy       8941 TTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGCGCCGGCGGGGGC 9000
Db       9281 TTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGCGCCGGCGGGGGC 9340

Qy       9001 GACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTTGGCCTACTCCTA 9060
Db       9341 GACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTTGGCCTACTCCTA 9400

Qy       9061 CTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCGG 9099
Db       9401 CTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCGG 9439
```

```
RESULT 4
AF177036
LOCUS       AF177036               9711 bp    RNA     linear   VRL 05-OCT-1999
DEFINITION  Hepatitis C virus clone pJ6CF, complete genome.
ACCESSION   AF177036
VERSION     AF177036.1  GI:6010579
KEYWORDS    .
SOURCE      Hepatitis C virus
  ORGANISM  Hepatitis C virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepacivirus.
REFERENCE   1  (bases 1 to 9711)
  AUTHORS   Yanagi,M., Purcell,R.H., Emerson,S.U. and Bukh,J.
  TITLE     Hepatitis C virus: an infectious molecular clone of a second major
            genotype (2a) and lack of viability of intertypic 1a and 2a
            chimeras
  JOURNAL   Virology 262 (1), 250-263 (1999)  sept. 15. 1999
  MEDLINE   99420396
   PUBMED   10489358
REFERENCE   2  (bases 1 to 9711)
  AUTHORS   Bukh,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-AUG-1999) Hepatitis Viruses Section, Laboratory of
            Infectious Diseases, National Institute of Allergy and Infectious
            Diseases, National Institutes of Health, Building 7, Room 201,
            7-Center Dr., Bethesda, MD 20892-0740, USA
FEATURES             Location/Qualifiers
     source          1. .9711
                     /organism="Hepatitis C virus"
                     /mol_type="genomic RNA"
                     /strain="HC-J6CH"
```

```
                     /db_xref="taxon:11103"
                     /clone="pJ6CF"
                     /note="infectious cDNA
                     genotype: 2a"
     CDS             341. .9442
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="AAF01178.1"
                     /db_xref="GI:6010580"
                     /translation="MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRL
                     GVRATRKTSERSQPRGRRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRG
                     SRPSWGPNDPRHRSRNVGKVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLED
                     GVNFATGNLPGCSFSIFLLALLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQA
                     AVLHVPGCVPCEKVGNASQCWIPVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALY
                     VGDLCGGVMLAAQMFIVSPQHHWFVQDCNCSIYPGTITGHRMAWDMMMNWSPTATMIL
                     AYAMRVPEVIIDIISGAHWGVMFGLAYFSMQGAWAKVVVILLLAAGVDARTHTVGGSA
                     AQTTGRLTSLFDMGPRQKIQLVNTNGSWHINRTALNCNDSLHTGFIASLFYTHSFNSS
                     GCPERMSACRSIEAFRVGWGALQYEDNVTNPEDMRPYCWHYPPRQCGVVSAKTVCGPV
                     YCFTPSPVVVGTTDRLGAPTYTWGENETDVFLLNSTRPPLGSWFGCTWMNSSGYTKTC
                     GAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLKCGSGPWLTPRCLIDYPYRLWHYPC
                     TVNYTIPKIRMYVGGVEHRLTAACNFTRGDRCNLEDRDRSQLSPLLHSTTEWAILPCS
                     YSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRWEWVILLFLLLADARVCACLW
                     MLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAWYIKGRVVPLATYSLTGLW
                     SFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYKTLLSRFLWWLCYLLTL
                     GEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVLGPAYLLKGALTRVP
                     YFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPMSDWAASGLRDLA
                     VAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPADGYTSKGWSLL
                     APITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISGVLWTVYHG
                     AGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRNADVIPA
                     RRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFIPVET
                     LDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNPSV
                     AATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD
                     ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIP
                     FYGRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDV
                     VVVATDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGR
                     TGRGRLGIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGL
                     PVCQDHLEFWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDV
                     MWKCLTRLKPTLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAG
                     GVLAAVAAYCLATGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQ
                     RIAEMLKSKIQGLLQQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLST
                     LPGNPAVASMMAFSAALTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAA
                     VGSIGLGKVLVDILAGYGAGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGV
                     ICAAILRRHVGPGEGAVQWMNRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTIT
                     SLLRRLHNWITEDCPIPCGGSWLRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQ
                     KGYKGVWAGTGIMTTRCPCGANISGNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQC
                     VPKPAPNFKVAIWRVAASEYAEVTQHGSYHYITGLTTDNLKVPCQLPSPEFFSWVDGV
                     QIHRFAPTPKPFFRDEVSFCVGLNSFVVGSQLPCDPEPDTDVLMSMLTDPSHITAETA
                     ARRLARGSPPSEASSSASQLSAPSLRATCTTHGKAYDVDMVDANLFMGGDVTRIESGS
                     KVVVLDSLDPMVEERSDLEPSIPSEYMLPKKRFPPALPAWARPDYNPPLVESWKRPDY
                     QPATVAGCALPPPRKTPTPPPRRRRTVGLSEDSIGDALQQLAIKSFGQPPPSGDSGLS
                     TGAGAADSGSQTPPDELALSETGSISSMPPLEGELGDPDLEPEQVEPQPPPQGGVAAP
                     GSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEEEKLPINPLSNSLLRYHNKVY
                     CTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIKLAASKVTARLLTMEEACQLTPPHS
                     ARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSETPIPTTIMAKNEVFCVDPTKGGK
                     KAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPAQRVEFLLKAWAEKK
                     DPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTERLYVGGPMFNSK
                     GQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPTMLVCGDDLVVISESQG
                     TEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQGRRRYYLTR
                     DPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDTLDQNLN
                     FEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLRAWKS
                     RARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGGDI
                     YHSVSRARPRLLLFGLLLLFVGVGLFLLPAR"
ORIGIN

  Query Match           100.0%;  Score 9099;  DB 14;  Length 9711;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 9099;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 400

Qy         61 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 460

Qy        121 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 520

Qy        181 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 580

Qy        241 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 640

Qy        301 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        641 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 700

Qy        361 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        701 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 760

Qy        421 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        761 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 820

Qy        481 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        821 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 880

Qy        541 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        881 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 940

Qy        601 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        941 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 1000

Qy        661 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1001 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 1060

Qy        721 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1061 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 1120

Qy        781 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1121 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 1180

Qy        841 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1181 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 1240

Qy        901 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1241 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 1300

Qy        961 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1301 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1360

Qy       1021 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1361 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1420

Qy       1081 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1421 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1480

Qy       1141 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1481 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1540

Qy       1201 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1541 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1600

Qy       1261 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1601 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1660

Qy       1321 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1661 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1720

Qy       1381 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1721 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1780

Qy       1441 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1781 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1840

Qy       1501 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1841 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1900

Qy       1561 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGGAGAATGAGACAGATGTCTTCCTA 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1901 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGGAGAATGAGACAGATGTCTTCCTA 1960

Qy       1621 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1961 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 2020

Qy       1681 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2021 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 2080

Qy       1741 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2081 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 2140

Qy       1801 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2141 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 2200

Qy       1861 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2201 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 2260

Qy       1921 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2261 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 2320

Qy       1981 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2321 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2380

Qy       2041 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2381 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2440

Qy       2101 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2441 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2500

Qy       2161 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2501 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2560

Qy       2221 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2561 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2620

Qy       2281 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATATTTTGTCATCTTTTTCGTGGCTGCTTGG 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2621 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATATTTTGTCATCTTTTTCGTGGCTGCTTGG 2680

Qy       2341 TACATCAAGGGTCGGGTAGTCCCCTTAGCTACCTATTCCCTCACTGGCCTGTGGTCCTTT 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2681 TACATCAAGGGTCGGGTAGTCCCCTTAGCTACCTATTCCCTCACTGGCCTGTGGTCCTTT 2740

Qy       2401 AGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTGCATGGC 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2741 AGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTGCATGGC 2800

Qy       2461 CAGATAGGAGCGGCTCTGCTGGTAATGATCACTCTCTTTACTCTCACCCCCGGGTATAAG 2520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2801 CAGATAGGAGCGGCTCTGCTGGTAATGATCACTCTCTTTACTCTCACCCCCGGGTATAAG 2860

Qy       2521 ACCCTTCTCAGCCGGTTTTTGTGGTGGTTGTGCTATCTTCTGACCCTGGGGGAAGCTATG 2580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2861 ACCCTTCTCAGCCGGTTTTTGTGGTGGTTGTGCTATCTTCTGACCCTGGGGGAAGCTATG 2920

Qy       2581 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGCATCATATGGGCC 2640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2921 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGCATCATATGGGCC 2980

Qy       2641 GTCGCCATATTCTACCCAGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 2700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2981 GTCGCCATATTCTACCCAGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 3040

Qy       2701 GGGCCTGCTTACCTCCTAAAAGGTGCTTTGACGCGCGTGCCGTACTTCGTCAGGGCTCAC 2760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3041 GGGCCTGCTTACCTCCTAAAAGGTGCTTTGACGCGCGTGCCGTACTTCGTCAGGGCTCAC 3100

Qy       2761 GCTCTACTGAGGATGTGCACCATGGCAAGGCATCTCGCGGGGGGCAGGTACGTCCAGATG 2820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3101 GCTCTACTGAGGATGTGCACCATGGCAAGGCATCTCGCGGGGGGCAGGTACGTCCAGATG 3160

Qy       2821 GCGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 2880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3161 GCGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 3220

Qy       2881 TCGGATTGGGCTGCTAGTGGCCTGCGGGACCTGGCGGTCGCCGTTGAGCCTATCATCTTC 2940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3221 TCGGATTGGGCTGCTAGTGGCCTGCGGGACCTGGCGGTCGCCGTTGAGCCTATCATCTTC 3280

Qy       2941 AGTCCGATGGAGAAGAAAGTCATTGTCTGGGGAGCGGAGACAGCTGCTTGTGGGGACATT 3000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3281 AGTCCGATGGAGAAGAAAGTCATTGTCTGGGGAGCGGAGACAGCTGCTTGTGGGGACATT 3340

Qy       3001 TTACACGGACTTCCCGTGTCCGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3341 TTACACGGACTTCCCGTGTCCGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3400

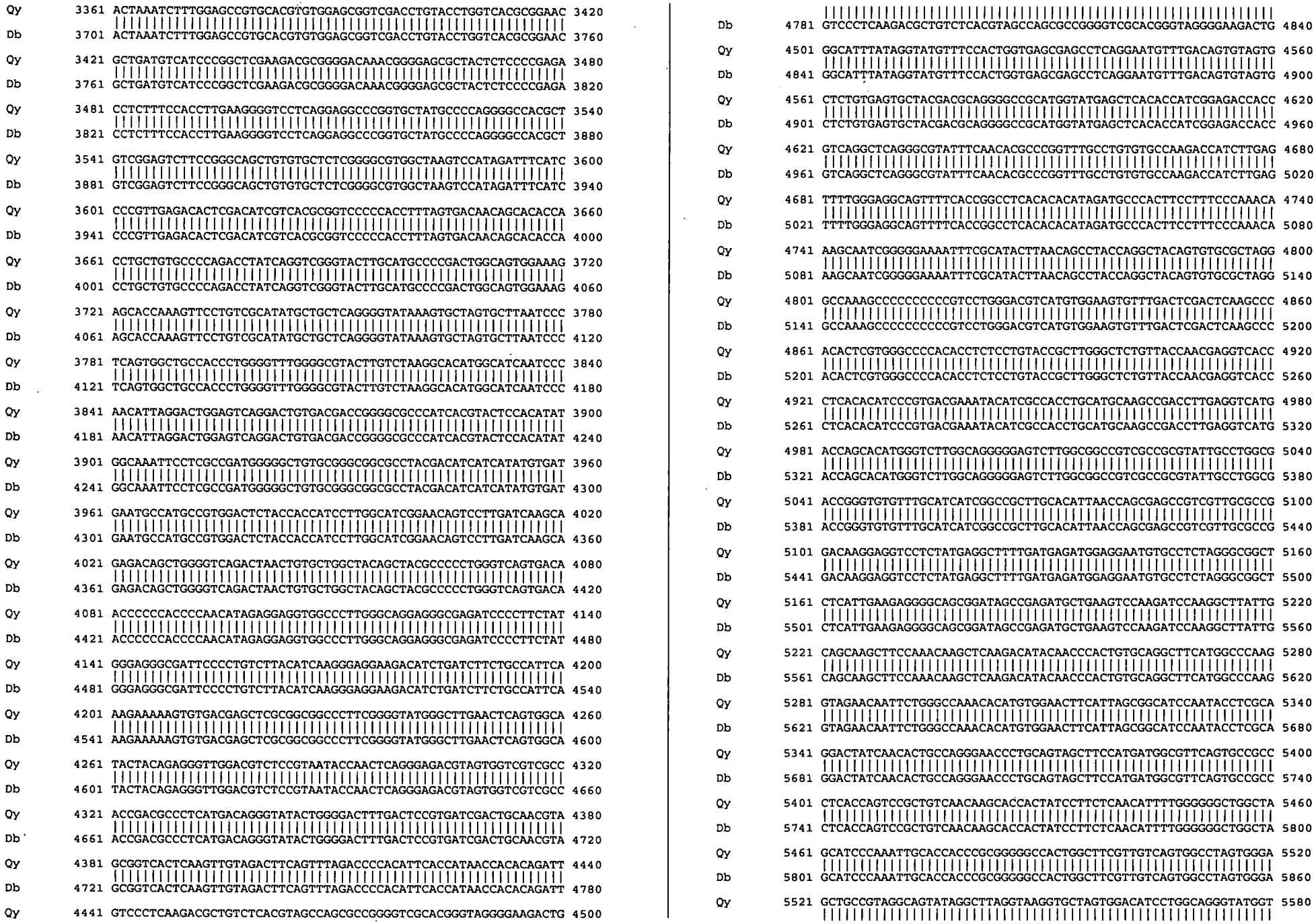
Qy       3061 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3401 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3460

Qy       3121 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCT 3180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3461 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCT 3520

Qy       3181 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3521 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3580

Qy       3241 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3581 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3640

Qy       3301 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGG 3360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3641 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGG 3700
```


Qy 3361 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3420
Db 3701 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3760
Qy 3421 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3480
Db 3761 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3820
Qy 3481 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3540
Db 3821 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3880
Qy 3541 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3600
Db 3881 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3940
Qy 3601 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 3660
Db 3941 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 4000
Qy 3661 CCTGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 3720
Db 4001 CCTGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 4060
Qy 3721 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 3780
Db 4061 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 4120
Qy 3781 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 3840
Db 4121 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 4180
Qy 3841 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 3900
Db 4181 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 4240
Qy 3901 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 3960
Db 4241 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 4300
Qy 3961 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4020
Db 4301 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4360
Qy 4021 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4080
Db 4361 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4420
Qy 4081 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4140
Db 4421 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4480
Qy 4141 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4200
Db 4481 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4540
Qy 4201 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4260
Db 4541 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4600
Qy 4261 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4320
Db 4601 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4660
Qy 4321 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4380
Db 4661 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4720
Qy 4381 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4440
Db 4721 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4780
Qy 4441 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4500
Db 4781 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4840
Qy 4501 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4560
Db 4841 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4900
Qy 4561 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4620
Db 4901 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4960
Qy 4621 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 4680
Db 4961 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 5020
Qy 4681 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 4740
Db 5021 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 5080
Qy 4741 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 4800
Db 5081 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 5140
Qy 4801 GCCAAAGCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 4860
Db 5141 GCCAAAGCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 5200
Qy 4861 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 4920
Db 5201 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 5260
Qy 4921 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 4980
Db 5261 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 5320
Qy 4981 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5040
Db 5321 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5380
Qy 5041 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5100
Db 5381 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5440
Qy 5101 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5160
Db 5441 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5500
Qy 5161 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5220
Db 5501 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5560
Qy 5221 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5280
Db 5561 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5620
Qy 5281 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5340
Db 5621 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5680
Qy 5341 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5400
Db 5681 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5740
Qy 5401 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5460
Db 5741 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5800
Qy 5461 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5520
Db 5801 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5860
Qy 5521 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5580

```
Db       5861 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5920

Qy       5581 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5921 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5980

Qy       5641 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 5700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5981 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 6040

Qy       5701 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 5760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6041 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 6100

Qy       5761 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 5820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6101 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 6160

Qy       5821 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 5880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6161 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 6220

Qy       5881 CTCAGAAGACTCCACAACTGGATTACTGAGGACTGCCCCATCCCATGCGGCGGCTCGTGG 5940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6221 CTCAGAAGACTCCACAACTGGATTACTGAGGACTGCCCCATCCCATGCGGCGGCTCGTGG 6280

Qy       5941 CTCCGCGATGTGTGGGACTGGGTTTGCACCATCCTAACAGACTTTAAAAATTGGCTGACC 6000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6281 CTCCGCGATGTGTGGGACTGGGTTTGCACCATCCTAACAGACTTTAAAAATTGGCTGACC 6340

Qy       6001 TCCAAATTATTCCCAAAGATGCCCGGCCTCCCCTTTGTCTCCTGTCAAAAGGGGTACAAG 6060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6341 TCCAAATTATTCCCAAAGATGCCCGGCCTCCCCTTTGTCTCCTGTCAAAAGGGGTACAAG 6400

Qy       6061 GGCGTGTGGGCCGGCACTGGCATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6401 GGCGTGTGGGCCGGCACTGGCATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6460

Qy       6121 GGCAATGTCCGCTTGGGCTCCATGAGAATCACGGGGCCTAAGACCTGCATGAATATCTGG 6180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6461 GGCAATGTCCGCTTGGGCTCCATGAGAATCACGGGGCCTAAGACCTGCATGAATATCTGG 6520

Qy       6181 CAGGGGACCTTTCCTATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCGCCA 6240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6521 CAGGGGACCTTTCCTATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCGCCA 6580

Qy       6241 AACTTTAAGGTCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6581 AACTTTAAGGTCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6640

Qy       6301 GGGTCATACCACTACATAACAGGACTCACCACTGATAACTTGAAAGTCCCCTGCCAACTA 6360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6641 GGGTCATACCACTACATAACAGGACTCACCACTGATAACTTGAAAGTCCCCTGCCAACTA 6700

Qy       6361 CCCTCTCCCGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCACA 6420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6701 CCCTCTCCCGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCACA 6760

Qy       6421 CCGAAGCCGTTTTTTCCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6761 CCGAAGCCGTTTTTTCCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6820

Qy       6481 GGGTCCCAGCTTCCTTGCGACCCTGAACCCGACACAGACGTATTGATGTCCATGCTAACA 6540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6821 GGGTCCCAGCTTCCTTGCGACCCTGAACCCGACACAGACGTATTGATGTCCATGCTAACA 6880

Qy       6541 GATCCATCTCATATCACGGCGGAGACTGCAGCGCGGCGTTTAGCGCGGGGGTCACCCCCA 6600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6881 GATCCATCTCATATCACGGCGGAGACTGCAGCGCGGCGTTTAGCGCGGGGGTCACCCCCA 6940

Qy       6601 TCCGAGGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 6660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6941 TCCGAGGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 7000

Qy       6661 ACCCACGGCAAAGCCTATGATGTGGACATGGTGGATGCTAACCTGTTCATGGGGGGCGAT 6720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7001 ACCCACGGCAAAGCCTATGATGTGGACATGGTGGATGCTAACCTGTTCATGGGGGGCGAT 7060

Qy       6721 GTGACTCGGATAGAGTCTGGGTCCAAAGTGGTCGTTCTGGACTCTCTCGACCCAATGGTC 6780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7061 GTGACTCGGATAGAGTCTGGGTCCAAAGTGGTCGTTCTGGACTCTCTCGACCCAATGGTC 7120

Qy       6781 GAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCAGAATACATGCTCCCCAAGAAGAGG 6840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7121 GAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCAGAATACATGCTCCCCAAGAAGAGG 7180

Qy       6841 TTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCGCTTGTGGAATCG 6900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7181 TTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCGCTTGTGGAATCG 7240

Qy       6901 TGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTCCCTCCTCCTAGG 6960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7241 TGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTCCCTCCTCCTAGG 7300

Qy       6961 AAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGTGAGGACTCCATA 7020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7301 AAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGTGAGGACTCCATA 7360

Qy       7021 GGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCCCCCAAGCGGCGAT 7080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7361 GGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCCCCCAAGCGGCGAT 7420

Qy       7081 TCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACGCCTCCTGATGAG 7140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7421 TCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACGCCTCCTGATGAG 7480

Qy       7141 TTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCTCGAGGGGGAGCTTGGA 7200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7481 TTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCTCGAGGGGGAGCTTGGA 7540

Qy       7201 GATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCCAGGGGGGGGTGGCA 7260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7541 GATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCAGGGGGGGGTGGCA 7600

Qy       7261 GCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGACGACTCCGTCGTG 7320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7601 GCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGACGACTCCGTCGTG 7660

Qy       7321 TGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGTAGTCCCGAAGAG 7380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7661 TGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGTAGTCCCGAAGAG 7720

Qy       7381 GAGAAGTTACCGATTAACCCCTTGAGCAACTCCCTGTTGCGATATCACAACAAGGTGTAC 7440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7721 GAGAAGTTACCGATTAACCCCTTGAGCAACTCCCTGTTGCGATATCACAACAAGGTGTAC 7780

Qy       7441 TGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTTGATAGGATGCAA 7500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7781 TGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTTGATAGGATGCAA 7840

Qy       7501 GTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCGGCCTCCAAGGTC 7560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7841 GTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCGGCCTCCAAGGTC 7900

Qy       7561 ACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCCCATTCTGCAAGA 7620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7901 ACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCCCATTCTGCAAGA 7960

Qy       7621 TCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGGGCCGTTAACCAC 7680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7961 TCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGGGCCGTTAACCAC 8020

Qy       7681 ATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATTCCCACAACCATT 7740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8021 ATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATTCCCACAACCATT 8080
```

```
Qy       7741 ATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAGAAAGCAGCTCGC 7800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8081 ATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAGAAAGCAGCTCGC 8140

Qy       7801 CTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCCCTTTATGACATT 7860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8141 CTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCCCTTTATGACATT 8200

Qy       7861 ACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAGTATTCCCCCGCT 7920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8201 ACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAGTATTCCCCCGCT 8260

Qy       7921 CAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCTATGGGTTTTTCG 7980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8261 CAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCTATGGGTTTTTCG 8320

Qy       7981 TATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGGACTGAGGAGTCC 8040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8321 TATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGGACTGAGGAGTCC 8380

Qy       8041 ATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACACTCGCTAACTGAG 8100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8381 ATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACACTCGCTAACTGAG 8440

Qy       8101 AGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGCGGGTACAGGCGT 8160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8441 AGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGCGGGTACAGGCGT 8500

Qy       8161 TGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACATGCTACGTGAAA 8220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8501 TGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACATGCTACGTGAAA 8560

Qy       8221 GCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTGGTATGCGGCGAT 8280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8561 GCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTGGTATGCGGCGAT 8620

Qy       8281 GACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGGAACCTGAGAGCC 8340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8621 GACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGGAACCTGAGAGCC 8680

Qy       8341 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCCAGACCGGAGTAT 8400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8681 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCCAGACCGGAGTAT 8740

Qy       8401 GATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTGGGCCCACAAGGC 8460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8741 GATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTGGGCCCACAAGGC 8800

Qy       8461 CGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGGGCTGCCTGGGAA 8520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8801 CGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGGGCTGCCTGGGAA 8860

Qy       8521 ACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAGTACGCCCCGACC 8580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8861 ACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAGTACGCCCCGACC 8920

Qy       8581 ATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATGGCTCAAGACACG 8640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8921 ATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATGGCTCAAGACACG 8980

Qy       8641 CTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTGAGTCCCTTGGAC 8700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8981 CTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTGAGTCCCTTGGAC 9040

Qy       8701 CTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTGCACACATACACT 8760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9041 CTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTGCACACATACACT 9100

Qy       8761 CCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGA 8820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9101 CCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGA 9160

Qy       8821 GCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGTGGGGGGAGAGCG 8880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9161 GCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGTGGGGGGAGAGCG 9220

Qy       8881 GCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 8940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9221 GCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 9280

Qy       8941 TTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGCGCCGGCGGGGGC 9000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9281 TTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGCGCCGGCGGGGGC 9340

Qy       9001 GACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTTGGCCTACTCCTA 9060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9341 GACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTTGGCCTACTCCTA 9400

Qy       9061 CTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCGG 9099
             ||||||||||||||||||||||||||||||||||||||||
Db       9401 CTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCGG 9439
```

```
RESULT 5
HPCPOLP
LOCUS       HPCPOLP                 9589 bp    RNA     linear   VRL 20-JUN-1998
DEFINITION  Hepatitis C virus genomic RNA for polyprotein, complete cds.
ACCESSION   D00944
VERSION     D00944.1  GI:221650
KEYWORDS    polyprotein.
SOURCE      Hepatitis C virus
  ORGANISM  Hepatitis C virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepacivirus.
REFERENCE   1  (bases 1 to 9589)
  AUTHORS   Okamoto,H., Okada,S., Sugiyama,Y., Kurai,K., Iizuka,H., Machida,A.,
            Miyakawa,Y. and Mayumi,M.
  TITLE     Nucleotide sequence of the genomic RNA of hepatitis C virus
            isolated from a human carrier: comparison with reported isolates
            for conserved and divergent regions
  JOURNAL   J. Gen. Virol. 72 (Pt 11), 2697-2704 (1991)
  MEDLINE   92044440
   PUBMED   1658196
REFERENCE   2  (sites)
  AUTHORS   Han,J.H. and Houghton,M.
  TITLE     Group specific sequences and conserved secondary structures at the
            3' end of HCV genome and its implication for viral replication
  JOURNAL   Nucleic Acids Res. 20 (13), 3520 (1992)
  MEDLINE   92335016
   PUBMED   1321416
REFERENCE   3  (sites)
  AUTHORS   Hotta,H., Doi,H., Hayashi,T., Purwanta,M., Soemarto,W.,
            Mizokami,M., Ohba,K. and Homma,M.
  TITLE     Analysis of the core and E1 envelope region sequences of a novel
            variant of hepatitis C virus obtained in Indonesia
  JOURNAL   Arch. Virol. 136 (1-2), 53-62 (1994)
  MEDLINE   94270990
   PUBMED   7545932
COMMENT     These data kindly submitted in computer readable form by:  Hiroaki
            Okamoto
            Immunology Division
            Jichi Medical School
            Kawachi-gun
            Tochigi-ken 329-04
            Japan
            Phone:  0285-44-2111 x3334
            Fax:    0285-44-1557.
FEATURES             Location/Qualifiers
     source          1. .9589
                     /organism="Hepatitis C virus"
                     /mol_type="genomic RNA"
                     /isolate="HC-J6"
                     /db_xref="taxon:11103"
     CDS             341. .9442
                     /codon_start=1
                     /product="polyprotein"
```

```
                     /protein_id="BAA00792.1"
                     /db_xref="GI:221651"
                     /translation="MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRL
                     GVRATRKTSERSQPRGRRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRG
                     SRPSWGPNDPRHRSRNVGKVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLED
                     GVNFATGNLPGCSFSIFLLALLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQA
                     AVLHVPGCVPCEKVGNTSRCWIPVSPNVAVQQPGALTQGLRTHIDMVVMSATLCSALY
                     VGDLCGGVMLAAQMFIVSPQHHWFVQDCNCSIYPGTITGHRMAWDMMMNWSPTATMIL
                     AYAMRVPEVIIDIIGGAHWGVMFGLAYFSMQGAWAKVVVILLLAAGVDAQTHTVGGST
                     AHNARTLTGMFSLGARQKIQLINTNGSWHINRTALNCNDSLHTGFLASLFYTHSFNSS
                     GCPERMSACRSIEAFRVGWGALQYEDNVTNPEDMRPYCWHYPPRQCGVVSASSVCGPV
                     YCFTPSPVVVGTTDRLGAPTYTWGENETDVFLLNSTRPPQGSWFGCTWMNSTGYTKTC
                     GAPPCRIRADFNASMDLLCPTDCFRKHPDTTYIKCGSGPWLTPRCLIDYPYRLWHYPC
                     TVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLEDRDRSQLSPLLHSTTEWAILPCT
                     YSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRWEWVVLLFLLLADARVCACLW
                     MLILLGQAEAALEKLVVLHAASAASCNGFLYFVIFFVAAWYIKGRVVPLATYSLTGLW
                     SFGLLLLALPQQAYAYDASVHGQIGAALLVLITLFTLTPGYKTLLSRFLWWLCYLLTL
                     AEAMVQEWAPPMQVRGGRDGIIWAVAIFCPGVVFDITKWLLAVLGPAYLLKGALTRVP
                     YFVRAHALLRMCTMVRHLAGGRYVQMVLLALGRWTGTYIYDHLTPMSDWAANGLRDLA
                     VAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPADGYTSKGWSLL
                     APITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTTISGVLWTVYHG
                     AGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRNADVIPA
                     RRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFIPVET
                     LDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNPSV
                     AATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD
                     ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIP
                     FYGRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNAVAYYRGLDVSVIPTQGDV
                     VVVATDALMTGFTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQTVPQDAVSRSQRRGR
                     TGRGRLGIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPAETTVRLRAYFNTPGL
                     PVCQDHLEFWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDV
                     MWKCLTRLKPTLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAG
                     GVLAAVAAYCLATGCVCIIGRLHVNQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQ
                     RIAEMLKSKIQGLLQQASKQAQDIQPAVQASWPKVEQFWAKHMWNFISGIQYLAGLST
                     LPGNPAVASMMAFSAALTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAA
                     VGSIGLGKVLVDILAGYGAGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGV
                     ICAAILRRHVGPGEGAVQWMNRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTIT
                     SLLRRLHNWITEDCPIPCSGSWLRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFISCQ
                     KGYKGVWAGTGIMTTRCPCGANISGNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQC
                     VPKPAPNFKIAIWRVAASEYAEVTQHGSYHYITGLTTDNLKVPCQLPSPEFFSWVDGV
                     QIHRFAPIPKPFFRDEVSFCVGLNSFVVGSQLPCDPEPDTDVLTSMLTDPSHITAETA
                     ARRLARGSPPSEASSSASQLSAPSLRATCTTHGKAYDVDMVDANLFMGGDVTRIESES
                     KVVVLDSLDPMVEERSDLEPSIPSEYMLPKKRFPPALPAWARPDYNPPLVESWKRPDY
                     QPATVAGCALPPPKKTPTPPPRRRRTVGLSESSIADALQQLAIKSFGQPPPSGDSGLS
                     TGADAADSGSRTPPDELALSETGSISSMPPLEGEPGDPDLEPEQVELQPPPQGGVVTP
                     GSGSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEEEKLPINPLSNSLLRYHNKVY
                     CTTSKSASLRAKKVTFDRMQALDAHYDSVLKDIKLAASKVTARLLTLEEACQLTPPHS
                     ARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDTQTPIPTTIMAKNEVFCVDPTKGGK
                     KAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPAQRVEFLLKAWAEKK
                     DPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTERLYVGGPMFNSK
                     GQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPTMLVCGDDLVVISESQG
                     TEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQGRRRYYLTR
                     DPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDTLDQNLN
                     FEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLRAWKS
                     RARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGGDI
                     YHSVSRARPRLLLLGLLLLFVGVGLFLLPAR"
ORIGIN

  Query Match          93.4%;  Score 8497.4;  DB 14;  Length 9589;
  Best Local Similarity   95.9%;  Pred. No. 0;
  Matches 8723;  Conservative    0;  Mismatches  376;  Indels    0;  Gaps    0;

Qy          1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 400

Qy         61 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 460

Qy        121 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 520

Qy        181 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 AGGCGCCAGCCCATCCCTAAGGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 580

Qy        241 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 640

Qy        301 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        641 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCCCGCAACGTGGGT 700

Qy        361 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        701 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTA 760

Qy        421 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        761 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 820

Qy        481 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        821 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 880

Qy        541 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        881 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 940

Qy        601 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 660
              ||||||||||| |||||||||||||||| |||||||||||||| ||||||||||||||||
Db        941 TACATGGTGACCAACGACTGCACCAATGATAGCATTACCTGGCAACTCCAGGCTGCTGTC 1000

Qy        661 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 720
              ||||||||||||||||||||||||||||||||||||||||||||||| |||||| ||||||
Db       1001 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATACATCTCGGTGCTGGATA 1060

Qy        721 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 780
              |||||||||||||||||||||||||||| |||||||||||||||||||||| ||||||
Db       1061 CCGGTCTCACCGAATGTGGCCGTGCAGCAGCCCGGCGCCCTCACGCAGGGCTTACGGACG 1120

Qy        781 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 840
              ||||| |||||||||||||||||||||||||||||||| || || ||||||||||||||
Db       1121 CACATTGACATGGTTGTGATGTCCGCCACGCTCTGCTCCGCTCTTTACGTGGGGGACCTC 1180

Qy        841 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 900
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1181 TGCGGTGGGGTGATGCTTGCAGCCCAGATGTTCATTGTCTCGCCACAGCACCACTGGTTT 1240

Qy        901 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 960
              || |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db       1241 GTGCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCGTGG 1300

Qy        961 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1020
              |||||||||||||||||||||||||||||||||||||| |||||||||||||||| |||
Db       1301 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCCTGGCGTACGCGATGCGCGTC 1360

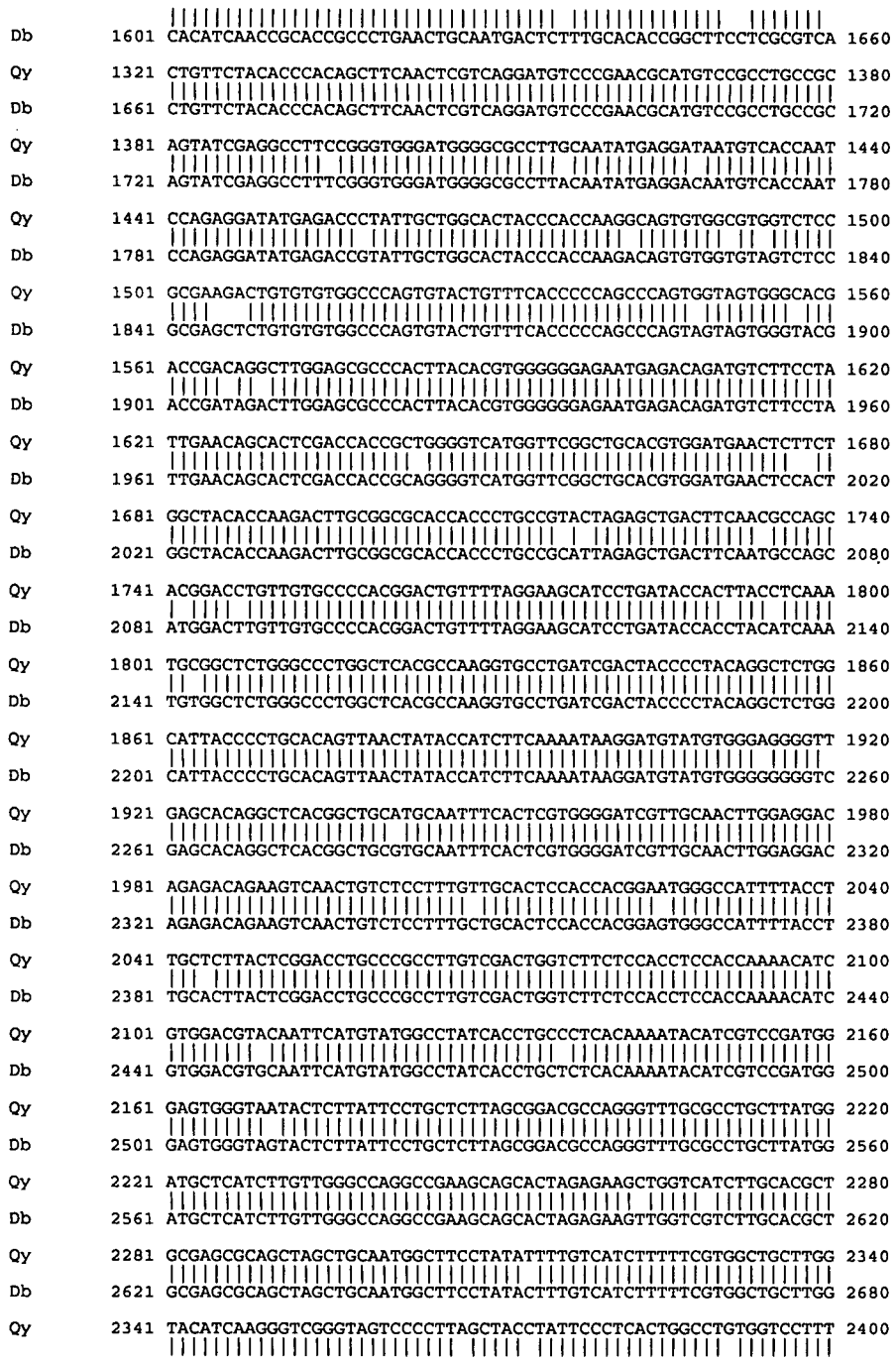
Qy       1021 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1080
              |||||||||||| |||||||||||||||||||||||||||||||||||||||| |||||
Db       1361 CCCGAGGTCATCATAGACATCATTGGCGGGGCTCATTGGGGCGTCATGTTCGGCTTAGCC 1420

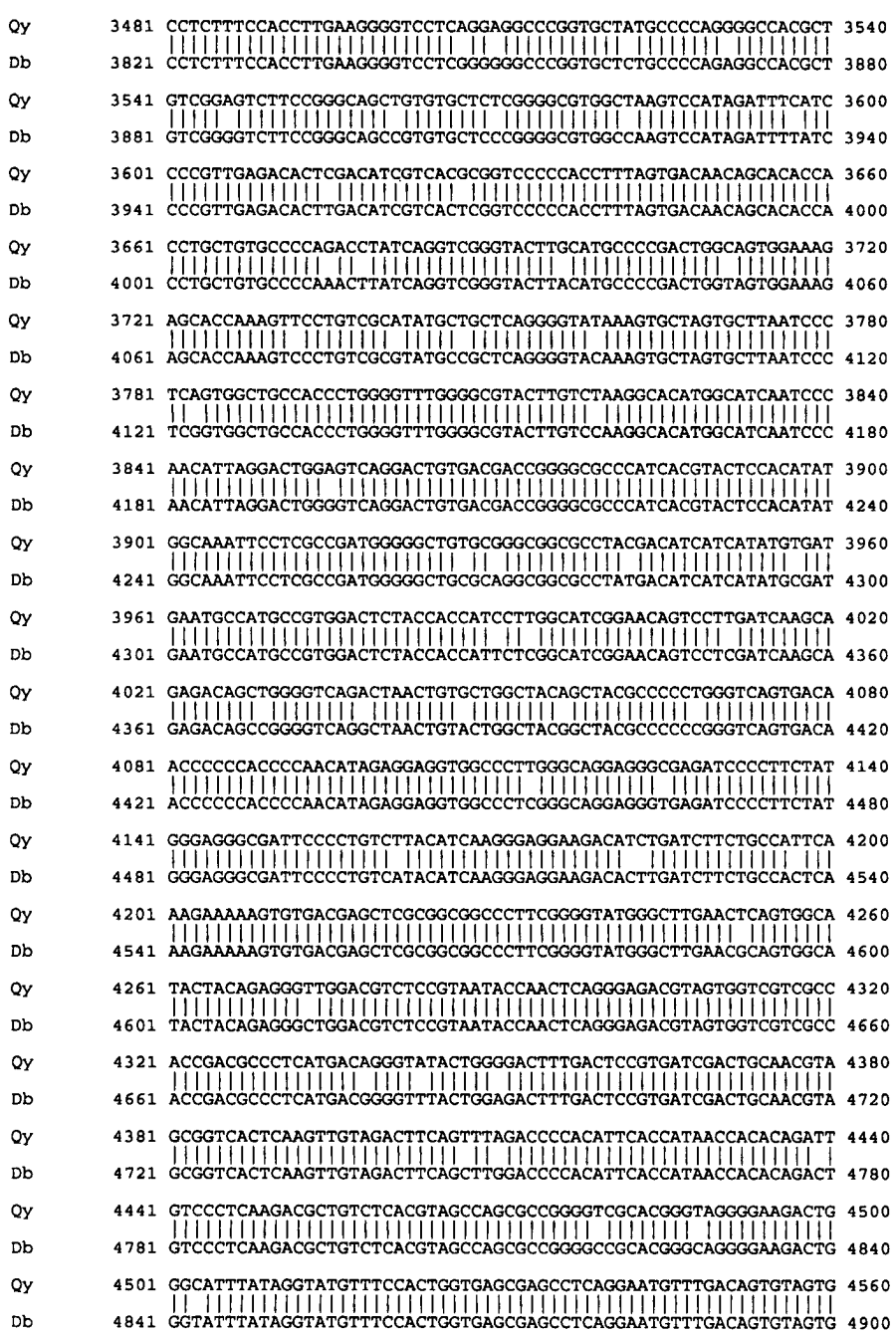
Qy       1081 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1140
              |||||||||||||||||||||||||||| |||||||||||||| ||| || ||||||||||
Db       1421 TACTTCTCTATGCAGGGAGCGTGGGCAAAAGTCGTTGTCATTCTTTTGCTGGCCGCCGGG 1480

Qy       1141 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1200
              |||||||||| |||||||| |||||||||||| ||||||| | | || || |||||||
Db       1481 GTGGACGCGCAAACCCATACCGTTGGGGGTTCTACCGCGCATAACGCCAGGACCCTCACC 1540

Qy       1201 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1260
              || | ||   | | ||  |||||||||||||||||| | |||||||||||||| |||
Db       1541 GGCATGTTCTCCCTTGGTGCCAGGCAGAAAATCCAGCTCATCAACACCAATGGCAGTTGG 1600

Qy       1261 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1320
```


Db 1601 CACATCAACCGCACCGCCCTGAACTGCAATGACTCTTTGCACACCGGCTTCCTCGCGTCA 1660
Qy 1321 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1380
Db 1661 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1720
Qy 1381 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1440
Db 1721 AGTATCGAGGCCTTTCGGGTGGGATGGGGCGCCTTACAATATGAGGACAATGTCACCAAT 1780
Qy 1441 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1500
Db 1781 CCAGAGGATATGAGACCGTATTGCTGGCACTACCCACCAAGACAGTGTGGTGTAGTCTCC 1840
Qy 1501 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1560
Db 1841 GCGAGCTCTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTAGTAGTGGGTACG 1900
Qy 1561 ACCGACAGGCTTGGAGCGCCCCACTTACACGTGGGGGGGAGAATGAGACAGATGTCTTCCTA 1620
Db 1901 ACCGATAGACTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1960
Qy 1621 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 1680
Db 1961 TTGAACAGCACTCGACCACCGCAGGGGTCATGGTTCGGCTGCACGTGGATGAACTCCACT 2020
Qy 1681 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 1740
Db 2021 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGCATTAGAGCTGACTTCAATGCCAGC 2080
Qy 1741 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 1800
Db 2081 ATGGACTTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACCTACATCAAA 2140
Qy 1801 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 1860
Db 2141 TGTGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 2200
Qy 1861 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 1920
Db 2201 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGGGGGGTC 2260
Qy 1921 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 1980
Db 2261 GAGCACAGGCTCACGGCTGCGTGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 2320
Qy 1981 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2040
Db 2321 AGAGACAGAAGTCAACTGTCTCCTTTGCTGCACTCCACCACGGAGTGGGCCATTTTACCT 2380
Qy 2041 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2100
Db 2381 TGCACTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2440
Qy 2101 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2160
Db 2441 GTGGACGTGCAATTCATGTATGGCCTATCACCTGCTCTCACAAAATACATCGTCCGATGG 2500
Qy 2161 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2220
Db 2501 GAGTGGGTAGTACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2560
Qy 2221 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2280
Db 2561 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGTTGGTCGTCTTGCACGCT 2620
Qy 2281 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATATTTTGTCATCTTTTTCGTGGCTGCTTGG 2340
Db 2621 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATACTTTGTCATCTTTTTCGTGGCTGCTTGG 2680
Qy 2341 TACATCAAGGGTCGGGTAGTCCCCTTAGCTACCTATTCCCTCACTGGCCTGTGGTCCTTT 2400
Db 2681 TACATCAAGGGTCGGGTAGTCCCCTTGGCTACTTATTCCCTCACTGGCCTATGGTCCTTT 2740
Qy 2401 AGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTGCATGGC 2460
Db 2741 GGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTACATGGT 2800
Qy 2461 CAGATAGGAGCGGCTCTGCTGGTAATGATCACTCTCTTTACTCTCACCCCCGGGTATAAG 2520
Db 2801 CAGATAGGAGCAGCTCTGTTGGTACTGATCACTCTCTTTACACTCACCCCCGGGTATAAG 2860
Qy 2521 ACCCTTCTCAGCCGGTTTTTGTGGTGGTTGTGCTATCTTCTGACCCTGGGGGAAGCTATG 2580
Db 2861 ACCCTTCTCAGCCGGTTTCTGTGGTGGTTGTGCTATCTTCTGACCCTGGCGGAAGCTATG 2920
Qy 2581 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGCATCATATGGGCC 2640
Db 2921 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGGATCATATGGGCC 2980
Qy 2641 GTCGCCATATTCTACCCAGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 2700
Db 2981 GTCGCCATATTCTGCCCGGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 3040
Qy 2701 GGGCCTGCTTACCTCCTAAAAGGTGCTTTGACGCGCGTGCCGTACTTCGTCAGGGCTCAC 2760
Db 3041 GGGCCTGCTTATCTCCTAAAAGGTGCTTTGACGCGTGTGCCGTACTTCGTCAGGGCTCAC 3100
Qy 2761 GCTCTACTGAGGATGTGCACCATGGCAAGGCATCTCGCGGGGGGCAGGTACGTCCAGATG 2820
Db 3101 GCTCTACTAAGGATGTGCACCATGGTAAGGCATCTCGCGGGGGGTAGGTACGTCCAGATG 3160
Qy 2821 GCGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 2880
Db 3161 GTGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 3220
Qy 2881 TCGGATTGGGCTGCTAGTGGCCTGCGGGACCTGGCGGTCGCCGTTGAGCCTATCATCTTC 2940
Db 3221 TCGGATTGGGCTGCTAATGGCCTGCGGGACTTGGCGGTCGCCGTGGAGCCTATCATCTTC 3280
Qy 2941 AGTCCGATGGAGAAGAAAGTCATTGTCTGGGGAGCGGAGACAGCTGCTTGTGGGGACATT 3000
Db 3281 AGTCCGATGGAGAAAAAAGTCATCGTCTGGGGAGCGGAGACAGCTGCTTGCGGGGATATC 3340
Qy 3001 TTACACGGACTTCCCGTGTCCGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3060
Db 3341 TTACACGGACTTCCCGTGTCCGCCCGACTTGGCCGGGAGGTCCTCCTTGGCCCAGCTGAT 3400
Qy 3061 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3120
Db 3401 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTATGCCCAGCAGACA 3460
Qy 3121 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCT 3180
Db 3461 CGCGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCC 3520
Qy 3181 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3240
Db 3521 GGGGAGATTCAGGTCCTGTCCACGGTCACTCAGTCCTTCCTCGGAACAACCATCTCGGGG 3580
Qy 3241 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3300
Db 3581 GTCTTATGGACTGTCTACCATGGAGCTGGCAACAAGACTCTAGCCGGCTCACGGGGTCCG 3640
Qy 3301 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGG 3360
Db 3641 GTCACACAGATGTACTCCAGTGCTGAGGGGGACTTAGTGGGGTGGCCCAGCCCCCCCGGG 3700
Qy 3361 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3420
Db 3701 ACCAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTATACCTGGTCACGCGAAAC 3760
Qy 3421 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3480
Db 3761 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAGCGAGGAGCGCTACTCTCCCCGAGA 3820


Qy 3481 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3540
Db 3821 CCTCTTTCCACCTTGAAGGGGTCCTCGGGGGGCCCGGTGCTCTGCCCCAGAGGCCACGCT 3880
Qy 3541 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3600
Db 3881 GTCGGGGTCTTCCGGGCAGCCGTGTGCTCCCGGGGCGTGGCCAAGTCCATAGATTTTATC 3940
Qy 3601 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 3660
Db 3941 CCCGTTGAGACACTTGACATCGTCACTCGGTCCCCCACCTTTAGTGACAACAGCACACCA 4000
Qy 3661 CCTGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 3720
Db 4001 CCTGCTGTGCCCCAAACTTATCAGGTCGGGTACTTACATGCCCCGACTGGTAGTGGAAAG 4060
Qy 3721 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 3780
Db 4061 AGCACCAAAGTCCCTGTCGCGTATGCCGCTCAGGGGTACAAAGTGCTAGTGCTTAATCCC 4120
Qy 3781 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 3840
Db 4121 TCGGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCCAAGGCACATGGCATCAATCCC 4180
Qy 3841 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 3900
Db 4181 AACATTAGGACTGGGGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 4240
Qy 3901 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 3960
Db 4241 GGCAAATTCCTCGCCGATGGGGGCTGCGCAGGCGGCGCCTATGACATCATCATATGCGAT 4300
Qy 3961 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4020
Db 4301 GAATGCCATGCCGTGGACTCTACCACCATTCTCGGCATCGGAACAGTCCTCGATCAAGCA 4360
Qy 4021 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4080
Db 4361 GAGACAGCCGGGGTCAGGCTAACTGTACTGGCTACGGCTACGCCCCCCGGGTCAGTGACA 4420
Qy 4081 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4140
Db 4421 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTCGGGCAGGAGGGTGAGATCCCCTTCTAT 4480
Qy 4141 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4200
Db 4481 GGGAGGGCGATTCCCCTGTCATACATCAAGGGAGGAAGACACTTGATCTTCTGCCACTCA 4540
Qy 4201 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4260
Db 4541 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACGCAGTGGCA 4600
Qy 4261 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4320
Db 4601 TACTACAGAGGGCTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4660
Qy 4321 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4380
Db 4661 ACCGACGCCCTCATGACGGGGTTTACTGGAGACTTTGACTCCGTGATCGACTGCAACGTA 4720
Qy 4381 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4440
Db 4721 GCGGTCACTCAAGTTGTAGACTTCAGCTTGGACCCCACATTCACCATAACCACACAGACT 4780
Qy 4441 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4500
Db 4781 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGCCGCACGGGCAGGGGAAGACTG 4840
Qy 4501 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4560
Db 4841 GGTATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4900
Qy 4561 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4620
Db 4901 CTCTGCGAGTGCTACGATGCAGGGGCCGCATGGTATGAGCTCACACCAGCGGAGACCACC 4960
Qy 4621 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 4680
Db 4961 GTCAGGCTCAGAGCATATTTCAACACACCTGGTTTGCCTGTGTGCCAAGACCATCTTGAG 5020
Qy 4681 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 4740
Db 5021 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 5080
Qy 4741 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 4800
Db 5081 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 5140
Qy 4801 GCCAAAGCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 4860
Db 5141 GCCAAAGCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 5200
Qy 4861 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 4920
Db 5201 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 5260
Qy 4921 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 4980
Db 5261 CTCACGCATCCTGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 5320
Qy 4981 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5040
Db 5321 ACCAGCACGTGGGTCTTAGCTGGGGGGGTCTTGGCGGCCGTCGCCGCGTACTGCCTGGCG 5380
Qy 5041 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5100
Db 5381 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACGTTAACCAGCGAGCCGTCGTTGCACCG 5440
Qy 5101 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5160
Db 5441 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGAGCGGCT 5500
Qy 5161 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5220
Db 5501 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5560
Qy 5221 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5280
Db 5561 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCGCTGTGCAGGCTTCTTGGCCCAAG 5620
Qy 5281 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5340
Db 5621 GTAGAGCAATTCTGGGCCAAACACATGTGGAACTTCATCAGCGGCATTCAATACCTCGCA 5680
Qy 5341 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5400
Db 5681 GGACTATCAACACTGCCAGGGAACCCTGCTGTAGCTTCCATGATGGCATTCAGTGCCGCC 5740
Qy 5401 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5460
Db 5741 CTCACCAGTCCGTTGTCAACTAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5800
Qy 5461 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5520
Db 5801 GCATCCCAAATTGCGCCTCCCGCGGGGGCTACCGGCTTCGTCGTCAGTGGCCTGGTGGGG 5860
Qy 5521 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5580
Db 5861 GCTGCCGTAGGCAGCATAGGCTTGGGTAAGGTGCTGGTGGACATCCTGGCAGGGTATGGT 5920
Qy 5581 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5640
Db 5921 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5980
Qy 5641 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 5700

```
Db       5981 GAGGATGTTGTCAACCTGCTGCCTGGAATTCTGTCTCCGGGTGCCCTGGTGGTGGGAGTC 6040
Qy       5701 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 5760
Db       6041 ATCTGCGCGGCCATCCTGCGCCGACACGTGGGACCGGGGGAAGGCGCTGTCCAATGGATG 6100
Qy       5761 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 5820
Db       6101 AATAGGCTCATTGCCTTTGCTTCCAGAGGAAACCACGTCGCCCCCACCCACTACGTGACG 6160
Qy       5821 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 5880
Db       6161 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 6220
Qy       5881 CTCAGAAGACTCCACAACTGGATTACTGAGGACTGCCCCATCCCATGCGGCGGCTCGTGG 5940
Db       6221 CTCAGGAGACTCCACAACTGGATTACTGAAGACTGCCCCATCCCATGCAGCGGCTCGTGG 6280
Qy       5941 CTCCGCGATGTGTGGGACTGGGTTTGCACCATCCTAACAGACTTTAAAAATTGGCTGACC 6000
Db       6281 CTCCGCGATGTGTGGGATTGGGTTTGCACCATCCTAACAGACTTTAAAAACTGGCTGACC 6340
Qy       6001 TCCAAATTATTCCCAAAGATGCCCGGCCTCCCCTTTGTCTCCTGTCAAAAGGGGTACAAG 6060
Db       6341 TCCAAATTGTTCCCAAAGATGCCTGGTCTCCCCTTTATCTCTTGTCAAAAGGGGTACAAG 6400
Qy       6061 GGCGTGTGGGCCGGCACTGGCATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6120
Db       6401 GGCGTGTGGGCTGGCACTGGTATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6460
Qy       6121 GGCAATGTCCGCTTGGGCTCCATGAGAATCACGGGGCCTAAGACCTGCATGAATATCTGG 6180
Db       6461 GGCAATGTCCGCCTGGGCTCCATGAGAATTACGGGGCCCAAAACCTGCATGAATATCTGG 6520
Qy       6181 CAGGGGACCTTTCCTATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCGCCA 6240
Db       6521 CAGGGGACCTTTCCCATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCACCA 6580
Qy       6241 AACTTTAAGGTCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6300
Db       6581 AACTTTAAGATCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6640
Qy       6301 GGGTCATACCACTACATAACAGGACTCACCACTGATAACTTGAAAGTCCCCTGCCAACTA 6360
Db       6641 GGGTCATACCACTACATAACAGGACTTACCACTGATAACTTGAAAGTTCCTTGCCAACTA 6700
Qy       6361 CCCTCTCCCGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCACA 6420
Db       6701 CCTTCTCCAGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCATA 6760
Qy       6421 CCGAAGCCGTTTTTTCCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6480
Db       6761 CCGAAGCCGTTTTTTTCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6820
Qy       6481 GGGTCCCAGCTTCCTTGCGACCCTGAACCCGACACAGACGTATTGATGTCCATGCTAACA 6540
Db       6821 GGGTCTCAGCTCCCTTGCGATCCTGAACCTGACACAGACGTATTGACGTCCATGCTAACA 6880
Qy       6541 GATCCATCTCATATCACGGCGGAGACTGCAGCGCGGCGTTTAGCGCGGGGGTCACCCCCA 6600
Db       6881 GACCCATCCCATATCACGGCGGAGACTGCAGCGCGGCGTTTGGCACGGGGGTCACCCCCG 6940
Qy       6601 TCCGAGGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 6660
Db       6941 TCCGAGGCAAGCTCCTCAGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 7000
Qy       6661 ACCCACGGCAAAGCCTATGATGTGGACATGGTGGATGCTAACCTGTTCATGGGGGGCGAT 6720
Db       7001 ACCCACGGCAAGGCCTATGATGTGGACATGGTGGATGCCAACCTGTTCATGGGGGGCGAT 7060
Qy       6721 GTGACTCGGATAGAGTCTGGGTCCAAAGTGGTCGTTCTGGACTCTCTCGACCCAATGGTC 6780
Db       7061 GTGACCCGGATAGAGTCTGAGTCCAAAGTGGTCGTTCTGGACTCTCTCGACCCAATGGTC 7120
Qy       6781 GAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCAGAATACATGCTCCCCAAGAAGAGG 6840
Db       7121 GAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCGGAATATATGCTCCCCAAGAAGAGA 7180
Qy       6841 TTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCGCTTGTGGAATCG 6900
Db       7181 TTCCCACCAGCCTTACCGGCTTGGGCACGGCCTGATTACAACCCACCGCTTGTGGAATCG 7240
Qy       6901 TGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTCCCTCCTCCTAGG 6960
Db       7241 TGGAAGAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGCGCTCTCCCCCCCCCTAAG 7300
Qy       6961 AAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGTGAGGACTCCATA 7020
Db       7301 AAAACCCCGACGCCTCCCCCAAGGAGACGCCGGACAGTGGGTCTGAGTGAGAGCTCCATA 7360
Qy       7021 GGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCCCCAAGCGGCGAT 7080
Db       7361 GCAGATGCCCTACAACAGCTGGCCATCAAGTCCTTTGGCCAGCCCCCCCCAAGCGGCGAT 7420
Qy       7081 TCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACGCCTCCTGATGAG 7140
Db       7421 TCAGGCCTTTCCACGGGGGCGGACGCAGCCGATTCCGGCAGTCGGACGCCCCCCGATGAG 7480
Qy       7141 TTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCTCGAGGGGGAGCTTGGA 7200
Db       7481 TTGGCCCTTTCGGAGACAGGTTCCATCTCCTCCATGCCCCCTCTCGAGGGGGAGCCTGGA 7540
Qy       7201 GATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCAGGGGGGGGTGGCA 7260
Db       7541 GATCCAGACTTGGAGCCTGAGCAGGTAGAGCTTCAACCTCCCCCCCAGGGGGGGGTGGTA 7600
Qy       7261 GCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGACGACTCCGTCGTG 7320
Db       7601 ACCCCCGGCTCAGGCTCGGGGTCTTGGTCTACTTGCTCCGAGGAGGACGACTCCGTCGTG 7660
Qy       7321 TGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGTAGTCCCGAAGAG 7380
Db       7661 TGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGTAGCCCCGAAGAG 7720
Qy       7381 GAGAAGTTACCGATTAACCCCTTGAGCAACTCCCTGTTGCGATATCACAACAAGGTGTAC 7440
Db       7721 GAAAAGTTGCCAATTAACCCCTTGAGCAACTCCCTGTTGCGATATCACAACAAGGTGTAC 7780
Qy       7441 TGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTTGATAGGATGCAA 7500
Db       7781 TGTACCACATCAAAGAGCGCCTCATTAAGGGCTAAAAAGGTAACTTTTGATAGGATGCAA 7840
Qy       7501 GTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCGGCCTCCAAGGTC 7560
Db       7841 GCGCTCGACGCTCATTATGACTCAGTCTTGAAGGACATTAAGCTAGCGGCCTCCAAGGTC 7900
Qy       7561 ACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCCCATTCTGCAAGA 7620
Db       7901 ACCGCAAGGCTTCTCACTTTAGAGGAGGCCTGCCAGTTAACTCCACCCCACTCTGCAAGA 7960
Qy       7621 TCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGGGCCGTTAACCAC 7680
Db       7961 TCCAAGTATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGAGCCGTTAACCAC 8020
Qy       7681 ATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATTCCCACAACCATT 7740
Db       8021 ATCAAGTCCGTGTGGAAGGACCTCCTGGAAGACACACAAACACCAATTCCTACAACCATC 8080
Qy       7741 ATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAGAAAGCAGCTCGC 7800
Db       8081 ATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGTAAGAAAGCAGCTCGC 8140
Qy       7801 CTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCCCTTTATGACATT 7860
Db       8141 CTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAAATGGCCCTTTATGATATC 8200
```

```
Qy       7861 ACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAGTATTCCCCCGCT 7920
             |||||||| |||||||||||||||||||||||||||||||||||||||||| ||||||||
Db       8201 ACACAAAAGCTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAGTACTCCCCCGCT 8260

Qy       7921 CAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCTATGGGTTTTTCG 7980
             |||||||| |||||||||||||| |||||||||||||||| |||||||||||||||||||
Db       8261 CAGCGGGTGGAGTTTCTCTTGAAGGCATGGGCGGAAAAGAAAGACCCTATGGGTTTTTCG 8320

Qy       7981 TATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGGACTGAGGAGTCC 8040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8321 TATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGGACTGAGGAGTCC 8380

Qy       8041 ATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACACTCGCTAACTGAG 8100
             |||||||||||| || |||||||||||||||||||||||||||||||| || ||||||
Db       8381 ATATATCGGGCTTGTTCCTTGCCCGAGGAGGCCCACACTGCCATACACTCACTGACTGAG 8440

Qy       8101 AGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGCGGGTACAGGCGT 8160
             |||||||||||||||||||| |||||||||||||||||||| ||||||||||||||||||
Db       8441 AGACTTTACGTGGGAGGGCCCATGTTCAACAGCAAGGGCCAGACCTGCGGGTACAGGCGT 8500

Qy       8161 TGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACATGCTACGTGAAA 8220
             |||||||||||||||||| |||||||||||||||||||||||||||||| ||||||
Db       8501 TGCCGCGCCAGCGGGGTGCTTACCACTAGCATGGGGAACACCATCACATGCTATGTGAAA 8560

Qy       8221 GCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTGGTATGCGGCGAT 8280
             |||||||||||| ||||| |||||||||||||| |||||||||||||||||||||||||||
Db       8561 GCCTTAGCGGCCTGTAAGGCTGCAGGGATAATTGCGCCCACAATGCTGGTATGCGGCGAT 8620

Qy       8281 GACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGGAACCTGAGAGCC 8340
             |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db       8621 GACTTGGTTGTCATCTCAGAGAGCCAGGGGACCGAGGAGGACGAGCGGAACCTGAGAGCC 8680

Qy       8341 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCCAGACCGGAGTAT 8400
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db       8681 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCCAGACCGGAATAT 8740

Qy       8401 GATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTGGGCCCACAAGGC 8460
             || ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db       8741 GACCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGTTGGGCCCACAAGGC 8800

Qy       8461 CGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGGGCTGCCTGGGAA 8520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8801 CGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGGGCTGCCTGGGAA 8860

Qy       8521 ACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAGTACGCCCCGACC 8580
             |||||||||||||||||||||||||||||| |||||||||||||||||||||| |||
Db       8861 ACAGTTAGACACTCCCCTGTCAATTCATGGCTAGGAAACATCATCCAGTACGCCCCAACC 8920

Qy       8581 ATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATGGCTCAAGACACG 8640
             |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||| ||
Db       8921 ATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATGGCCCAAGATACT 8980

Qy       8641 CTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTGAGTCCCTTGGAC 8700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8981 CTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTGAGTCCCTTGGAC 9040

Qy       8701 CTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTGCACACATACACT 8760
             |||||||| |||||||||||||||| |||||||||||| |||||||||||||||||||||
Db       9041 CTCCCAGCCATAATTGAAAGGTTACACGGGCTTGACGCTTTCTCTCTGCACACATACACT 9100

Qy       8761 CCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGA 8820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9101 CCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGA 9160

Qy       8821 GCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGTGGGGGGAGAGCG 8880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9161 GCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGTGGGGGGAGAGCG 9220

Qy       8881 GCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 8940
             |||||||||||| |||||||||||||| ||||||||||||||||||||||||||||||||
Db       9221 GCCGTTTGCGGCCGATATCTCTTCAACTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 9280

Qy       8941 TTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGCGCCGGCGGGGGC 9000
             |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db       9281 TTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACTGTCGGCGCCGGCGGGGGC 9340

Qy       9001 GACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTTGGCCTACTCCTA 9060
             |||||||||||||||||||||||||||||||||||||||||||||||||||| ||| ||||||||||||||
Db       9341 GACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTACTCCTTGGCCTACTCCTA 9400

Qy       9061 CTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCGG 9099
             |||||||||||||||||||||||||||||||||||||||
Db       9401 CTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCGG 9439


RESULT 6
E07361
LOCUS       E07361                  9589 bp    RNA     linear   PAT 29-SEP-1997
DEFINITION  gRNA of Hepatitis non-A non-B virus,HCV-J6.
ACCESSION   E07361
VERSION     E07361.1  GI:5708558
KEYWORDS    JP 1994121689-A/1.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 9589)
  AUTHORS   Okamoto,H. and Nakamura,T.
  TITLE     NON-A NON-B HEPATITIC VIRUS GENE, POLYNUCLEOTIDE, POLYPEPTIDE,
            ANTIGEN AND ANTIBODY DETECTION SYSTEM
  JOURNAL   Patent: JP 1994121689-A 1 06-MAY-1994;
            NAKAMURA TETSUO
COMMENT     OS   Hepatitis non-A non-B virus
            PN   JP 1994121689-A/1
            PD   06-MAY-1994
            PF   09-AUG-1991 JP 1991287402
            PI   OKAMOTO HIROAKI, NAKAMURA TETSUO
            PC   C12N15/51,A61K39/29,A61K39/395,A61K39/395,C07K13/00,C12P21/02,
            PC   C12P21/08,
            PC   G01N33/53,G01N33/576,G01N33/577//A61B10/00;
            CC   strandedness: Single;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            FH   Key             Location/Qualifiers
            FH
            FT   source          1. .9589
            FT                   /organism='Hepatitis non-A non-B virus' FT
                         /strain='HCV-J6'.
FEATURES             Location/Qualifiers
     source          1. .9589
                     /organism="unidentified"
                     /mol_type="genomic RNA"
                     /db_xref="taxon:32644"
ORIGIN

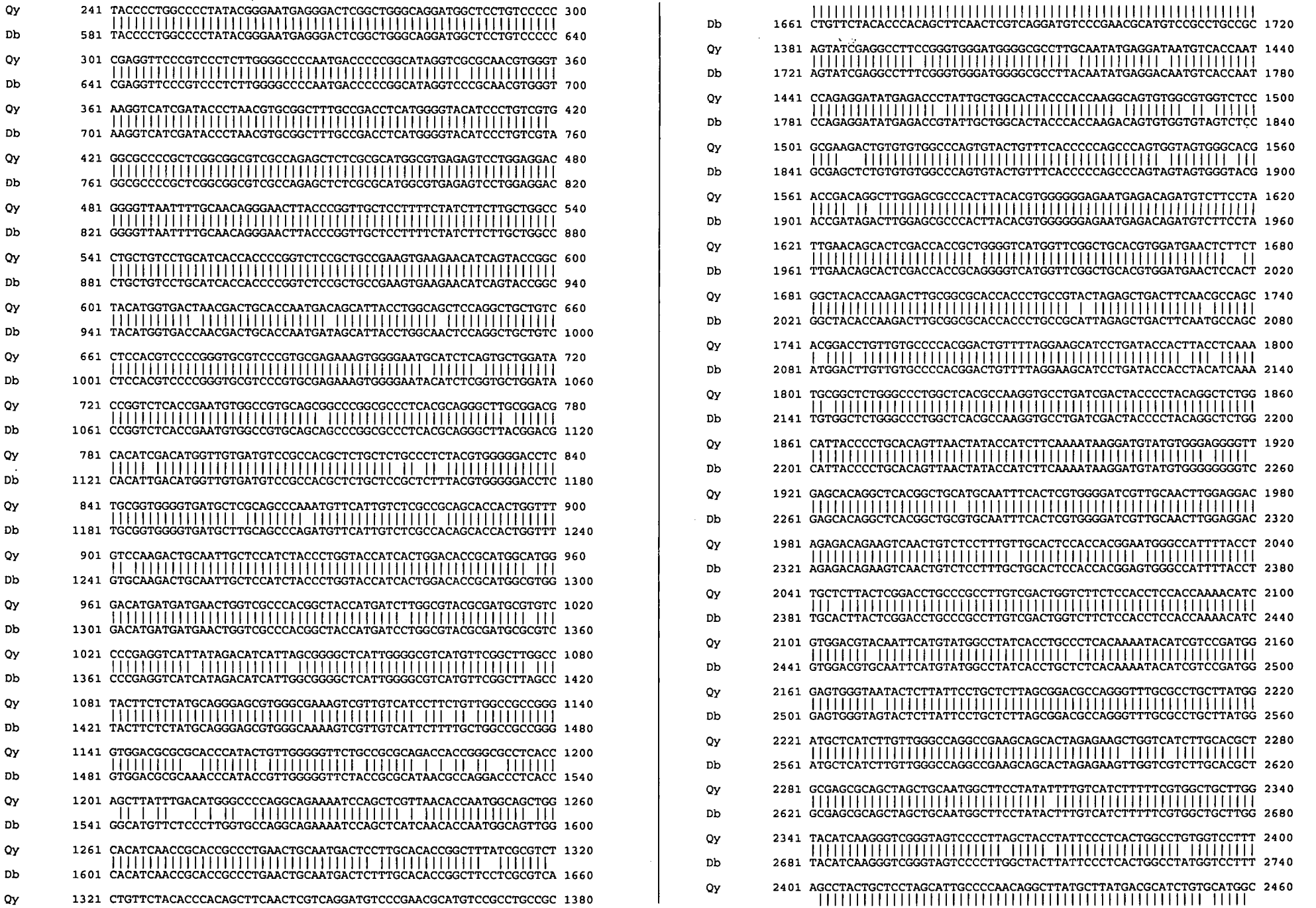
  Query Match             93.4%;  Score 8494.2;  DB 6;  Length 9589;
  Best Local Similarity   95.8%;  Pred. No. 0;
  Matches 8721;  Conservative    0;  Mismatches  378;  Indels    0;  Gaps    0;

Qy          1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 400

Qy         61 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 460

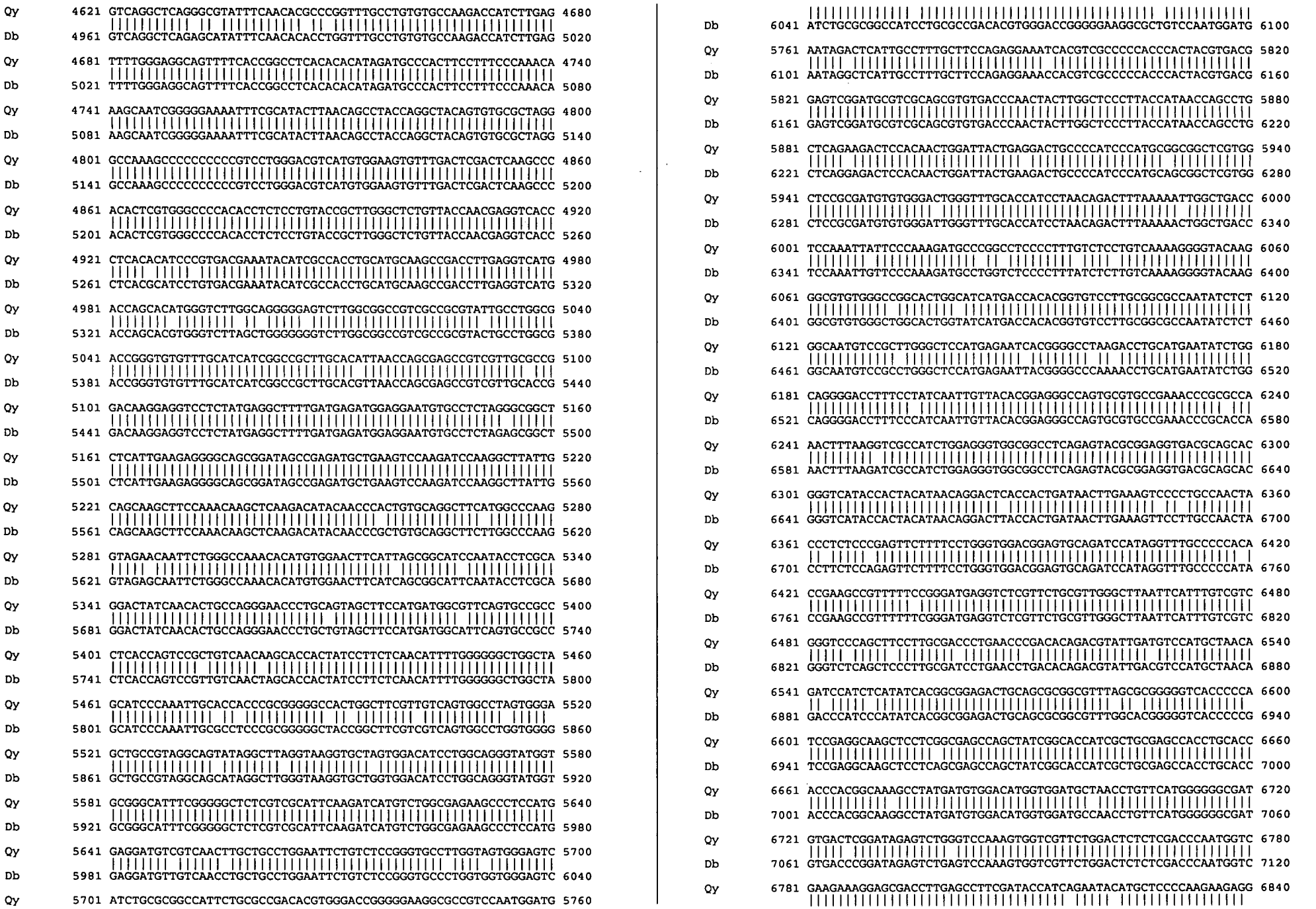
Qy        121 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 520

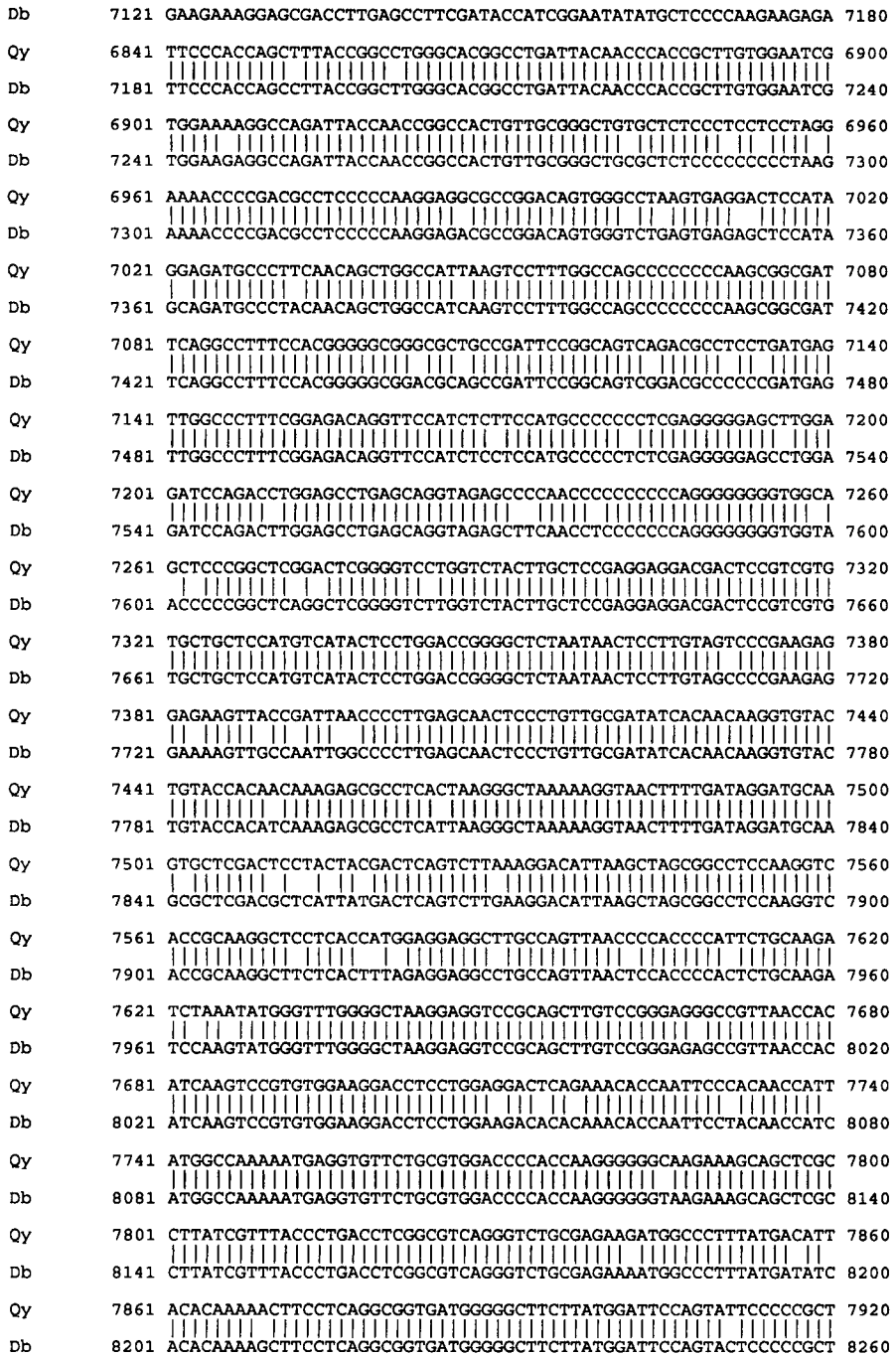
Qy        181 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 240
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        521 AGGCGCCAGCCCATCCCTAAGGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 580
```



```
Db       2741 GGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTACATGGT 2800
Qy       2461 CAGATAGGAGCGGCTCTGCTGGTAATGATCACTCTCTTTACTCTCACCCCCGGGTATAAG 2520
Db       2801 CAGATAGGAGCAGCTCTGTTGGTACTGATCACTCTCTTTACACTCACCCCCGGGTATAAG 2860
Qy       2521 ACCCTTCTCAGCCGGTTTTTGTGGTGGTTGTGCTATCTTCTGACCCTGGGGGAAGCTATG 2580
Db       2861 ACCCTTCTCAGCCGGTTTCTGTGGTGGTTGTGCTATCTTCTGACCCTGGCGGAAGCTATG 2920
Qy       2581 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGCATCATATGGGCC 2640
Db       2921 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGGATCATATGGGCC 2980
Qy       2641 GTCGCCATATTCTACCCAGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 2700
Db       2981 GTCGCCATATTCTGCCCGGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 3040
Qy       2701 GGGCCTGCTTACCTCCTAAAAGGTGCTTTGACGCGCGTGCCGTACTTCGTCAGGGCTCAC 2760
Db       3041 GGGCCTGCTTATCTCCTAAAAGGTGCTTTGACGCGTGTGCCGTACTTCGTCAGGGCTCAC 3100
Qy       2761 GCTCTACTGAGGATGTGCACCATGGCAAGGCATCTCGCGGGGGGCAGGTACGTCCAGATG 2820
Db       3101 GCTCTACTAAGGATGTGCACCATGGTAAGGCATCTCGCGGGGGGTAGGTACGTCCAGATG 3160
Qy       2821 GCGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 2880
Db       3161 GTGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 3220
Qy       2881 TCGGATTGGGCTGCTAGTGGCCTGCGGGACCTGGCGGTCGCCGTTGAGCCTATCATCTTC 2940
Db       3221 TCGGATTGGGCTGCTAATGGCCTGCGGGACTTGGCGGTCGCCGTGGAGCCTATCATCTTC 3280
Qy       2941 AGTCCGATGGAGAAGAAAGTCATTGTCTGGGGAGCGGAGACAGCTGCTTGTGGGGACATT 3000
Db       3281 AGTCCGATGGAGAAAAAAGTCATCGTCTGGGGAGCGGAGACAGCTGCTTGCGGGGATATC 3340
Qy       3001 TTACACGGACTTCCCGTGTCCGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3060
Db       3341 TTACACGGACTTCCCGTGTCCGCCCGACTTGGCCGGGAGGTCCTCCTTGGCCCAGCTGAT 3400
Qy       3061 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3120
Db       3401 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTATGCCCAGCAGACA 3460
Qy       3121 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCT 3180
Db       3461 CGCGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCC 3520
Qy       3181 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3240
Db       3521 GGGGAGATTCAGGTCCTGTCCACGGTCACTCAGTCCTTCCTCGGAACAACCATCTCGGGG 3580
Qy       3241 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3300
Db       3581 GTCTTATGGACTGTCTACCATGGAGCTGGCAACAAGACTCTAGCCGGCTCACGGGGTCCG 3640
Qy       3301 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGG 3360
Db       3641 GTCACACAGATGTACTCCAGTGCTGAGGGGGACTTAGTGGGGTGGCCCAGCCCCCCCGGG 3700
Qy       3361 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3420
Db       3701 ACCAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTATACCTGGTCACGCGAAAC 3760
Qy       3421 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3480
Db       3761 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAGCGAGGAGCGCTACTCTCCCCGAGA 3820
Qy       3481 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3540
Db       3821 CCTCTTTCCACCTTGAAGGGGTCCTCGGGGGGCCCGGTGCTCTGCCCCAGAGGCCACGCT 3880

Qy       3541 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3600
Db       3881 GTCGGGGTCTTCCGGGCAGCCGTGTGCTCCCGGGGCGTGGCCAAGTCCATAGATTTTATC 3940
Qy       3601 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 3660
Db       3941 CCCGTTGAGACACTTGACATCGTCACTCGGTCCCCCACCTTTAGTGACAACAGCACACCA 4000
Qy       3661 CCTGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 3720
Db       4001 CCTGCTGTGCCCCAAACTTATCAGGTCGGGTACTTACATGCCCCGACTGGTAGTGGAAAG 4060
Qy       3721 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 3780
Db       4061 AGCACCAAAGTCCCTGTCGCGTATGCCGCTCAGGGGTACAAAGTGCTAGTGCTTAATCCC 4120
Qy       3781 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 3840
Db       4121 TCGGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCCAAGGCACATGGCATCAATCCC 4180
Qy       3841 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 3900
Db       4181 AACATTAGGACTGGGGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 4240
Qy       3901 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 3960
Db       4241 GGCAAATTCCTCGCCGATGGGGGCTGCGCAGGCGGCGCCTATGACATCATCATATGCGAT 4300
Qy       3961 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4020
Db       4301 GAATGCCATGCCGTGGACTCTACCACCATTCTCGGCATCGGAACAGTCCTCGATCAAGCA 4360
Qy       4021 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4080
Db       4361 GAGACAGCCGGGGTCAGGCTAACTGTACTGGCTACGGCTACGCCCCCCGGGTCAGTGACA 4420
Qy       4081 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4140
Db       4421 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTCGGGCAGGAGGGTGAGATCCCCTTCTAT 4480
Qy       4141 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4200
Db       4481 GGGAGGGCGATTCCCCTGTCATACATCAAGGGAGGAAGACACTTGATCTTCTGCCACTCA 4540
Qy       4201 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4260
Db       4541 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACGCAGTGGCA 4600
Qy       4261 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4320
Db       4601 TACTACAGAGGGCTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4660
Qy       4321 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4380
Db       4661 ACCGACGCCCTCATGACGGGGTTTACTGGAGACTTTGACTCCGTGATCGACTGCAACGTA 4720
Qy       4381 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4440
Db       4721 GCGGTCACTCAAGTTGTAGACTTCAGCTTGGACCCCACATTCACCATAACCACACAGACT 4780
Qy       4441 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4500
Db       4781 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGCCGCACGGGCAGGGGAAGACTG 4840
Qy       4501 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4560
Db       4841 GGTATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4900
Qy       4561 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4620
Db       4901 CTCTGCGAGTGCTACGATGCAGGGGCCGCATGGTATGAGCTCACACCAGCGGAGACCACC 4960
```




Db 7121 GAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCGGAATATATGCTCCCCAAGAAGAGA 7180
Qy 6841 TTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCGCTTGTGGAATCG 6900
Db 7181 TTCCCACCAGCCTTACCGGCTTGGGCACGGCCTGATTACAACCCACCGCTTGTGGAATCG 7240
Qy 6901 TGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTCCCTCCTCCTAGG 6960
Db 7241 TGGAAGAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGCGCTCTCCCCCCCCCTAAG 7300
Qy 6961 AAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGTGAGGACTCCATA 7020
Db 7301 AAAACCCCGACGCCTCCCCCAAGGAGACGCCGGACAGTGGGTCTGAGTGAGAGCTCCATA 7360
Qy 7021 GGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCCCCCAAGCGGCGAT 7080
Db 7361 GCAGATGCCCTACAACAGCTGGCCATCAAGTCCTTTGGCCAGCCCCCCCCAAGCGGCGAT 7420
Qy 7081 TCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACGCCTCCTGATGAG 7140
Db 7421 TCAGGCCTTTCCACGGGGGCGGACGCAGCCGATTCCGGCAGTCGGACGCCCCCCGATGAG 7480
Qy 7141 TTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCTCGAGGGGGAGCTTGGA 7200
Db 7481 TTGGCCCTTTCGGAGACAGGTTCCATCTCCTCCATGCCCCCTCTCGAGGGGGAGCCTGGA 7540
Qy 7201 GATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCCAGGGGGGGGTGGCA 7260
Db 7541 GATCCAGACTTGGAGCCTGAGCAGGTAGAGCTTCAACCTCCCCCCCAGGGGGGGGTGGTA 7600
Qy 7261 GCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGACGACTCCGTCGTG 7320
Db 7601 ACCCCCGGCTCAGGCTCGGGGTCTTGGTCTACTTGCTCCGAGGAGGACGACTCCGTCGTG 7660
Qy 7321 TGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGTAGTCCCGAAGAG 7380
Db 7661 TGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGTAGCCCCGAAGAG 7720
Qy 7381 GAGAAGTTACCGATTAACCCCTTGAGCAACTCCCTGTTGCGATATCACAACAAGGTGTAC 7440
Db 7721 GAAAAGTTGCCAATTGGCCCCTTGAGCAACTCCCTGTTGCGATATCACAACAAGGTGTAC 7780
Qy 7441 TGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTTGATAGGATGCAA 7500
Db 7781 TGTACCACATCAAAGAGCGCCTCATTAAGGGCTAAAAAGGTAACTTTTGATAGGATGCAA 7840
Qy 7501 GTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCGGCCTCCAAGGTC 7560
Db 7841 GCGCTCGACGCTCATTATGACTCAGTCTTGAAGGACATTAAGCTAGCGGCCTCCAAGGTC 7900
Qy 7561 ACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCCCATTCTGCAAGA 7620
Db 7901 ACCGCAAGGCTTCTCACTTTAGAGGAGGCCTGCCAGTTAACTCCACCCCACTCTGCAAGA 7960
Qy 7621 TCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGGGCCGTTAACCAC 7680
Db 7961 TCCAAGTATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGAGCCGTTAACCAC 8020
Qy 7681 ATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATTCCCACAACCATT 7740
Db 8021 ATCAAGTCCGTGTGGAAGGACCTCCTGGAAGACACACAAACACCAATTCCTACAACCATC 8080
Qy 7741 ATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAGAAAGCAGCTCGC 7800
Db 8081 ATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGTAAGAAAGCAGCTCGC 8140
Qy 7801 CTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCCCTTTATGACATT 7860
Db 8141 CTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAAATGGCCCTTTATGATATC 8200
Qy 7861 ACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAGTATTCCCCCGCT 7920
Db 8201 ACACAAAAGCTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAGTACTCCCCCGCT 8260
Qy 7921 CAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCTATGGGTTTTTCG 7980
Db 8261 CAGCGGGTGGAGTTTCTCTTGAAGGCATGGGCGGAAAAGAAAGACCCTATGGGTTTTTCG 8320
Qy 7981 TATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGGACTGAGGAGTCC 8040
Db 8321 TATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGGACTGAGGAGTCC 8380
Qy 8041 ATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACACTCGCTAACTGAG 8100
Db 8381 ATATATCGGGCTTGTTCCTTGCCCGAGGAGGCCCACACTGCCATACACTCACTGACTGAG 8440
Qy 8101 AGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGCGGGTACAGGCGT 8160
Db 8441 AGACTTTACGTGGGAGGGCCCATGTTCAACAGCAAGGGCCAGACCTGCGGGTACAGGCGT 8500
Qy 8161 TGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACATGCTACGTGAAA 8220
Db 8501 TGCCGCGCCAGCGGGGTGCTTACCACTAGCATGGGGAACACCATCACATGCTATGTGAAA 8560
Qy 8221 GCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTGGTATGCGGCGAT 8280
Db 8561 GCCTTAGCGGCCTGTAAGGCTGCAGGGATAATTGCGCCCACAATGCTGGTATGCGGCGAT 8620
Qy 8281 GACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGGAACCTGAGAGCC 8340
Db 8621 GACTTGGTTGTCATCTCAGAGAGCCAGGGGACCGAGGAGGACGAGCGGAACCTGAGAGCC 8680
Qy 8341 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCCAGACCGGAGTAT 8400
Db 8681 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCCAGACCGGAATAT 8740
Qy 8401 GATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTGGGCCCACAAGGC 8460
Db 8741 GACCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGTTGGGCCCACAAGGC 8800
Qy 8461 CGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGGGCTGCCTGGGAA 8520
Db 8801 CGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGGGCTGCCTGGGAA 8860
Qy 8521 ACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAGTACGCCCCGACC 8580
Db 8861 ACAGTTAGACACTCCCCTGTCAATTCATGGCTAGGAAACATCATCCAGTACGCCCCAACC 8920
Qy 8581 ATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATGGCTCAAGACACG 8640
Db 8921 ATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATGGCCCAAGATACT 8980
Qy 8641 CTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTGAGTCCCTTGGAC 8700
Db 8981 CTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTGAGTCCCTTGGAC 9040
Qy 8701 CTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTGCACACATACACT 8760
Db 9041 CTCCCAGCCATAATTGAAAGGTTACACGGGCTTGACGCTTTCTCTCTGCACACATACACT 9100
Qy 8761 CCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGA 8820
Db 9101 CCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGA 9160
Qy 8821 GCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGTGGGGGGAGAGCG 8880
Db 9161 GCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGTGGGGGGAGAGCG 9220
Qy 8881 GCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 8940
Db 9221 GCCGTTTGCGGCCGATATCTCTTCAACTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 9280
Qy 8941 TTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGCGCCGGCGGGGGC 9000
Db 9281 TTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACTGTCGGCGCCGGCGGGGGC 9340

```
Qy       9001 GACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTTGGCCTACTCCTA 9060
             ||||||||||||||||||||||||||||||||||||||||||| ||| ||||||||||||||
Db       9341 GACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTACTCCTTGGCCTACTCCTA 9400

Qy       9061 CTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCGG 9099
             |||||||||||||||||||||||||||||||||||||||
Db       9401 CTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCGG 9439


RESULT 7
E07362
LOCUS       E07362                  9589 bp    RNA     linear   PAT 29-SEP-1997
DEFINITION  cDNA of Hepatitis non-A non-B virus.
ACCESSION   E07362
VERSION     E07362.1  GI:2175501
KEYWORDS    JP 1994121689-A/2.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 9589)
  AUTHORS   Okamoto,H. and Nakamura,T.
  TITLE     NON-A NON-B HEPATITIC VIRUS GENE, POLYNUCLEOTIDE, POLYPEPTIDE,
            ANTIGEN AND ANTIBODY DETECTION SYSTEM
  JOURNAL   Patent: JP 1994121689-A 2 06-MAY-1994;
            NAKAMURA TETSUO
COMMENT     OS   Hepatitis non-A non-B virus
            PN   JP 1994121689-A/2
            PD   06-MAY-1994
            PF   09-AUG-1991 JP 1991287402
            PI   OKAMOTO HIROAKI, NAKAMURA TETSUO
            PC   C12N15/51,A61K39/29,A61K39/395,A61K39/395,C07K13/00,C12P21/02,
            PC   C12P21/08,
            PC   G01N33/53,G01N33/576,G01N33/577//A61B10/00;
            CC   strandedness: Single;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            FH   Key             Location/Qualifiers
            FH
            FT   source          1. .9589
            FT                   /organism='Hepatitis non-A non-B virus' FT
                        /strain='HCV-J6'.
FEATURES             Location/Qualifiers
     source          1. .9589
                     /organism="unidentified"
                     /mol_type="genomic RNA"
                     /db_xref="taxon:32644"
ORIGIN

  Query Match             93.4%;  Score 8494.2;  DB 6;  Length 9589;
  Best Local Similarity   95.8%;  Pred. No. 0;
  Matches 8721;  Conservative    0;  Mismatches  378;  Indels    0;  Gaps    0;

Qy          1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 400

Qy         61 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 460

Qy        121 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 520

Qy        181 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 240
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        521 AGGCGCCAGCCCATCCCTAAGGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 580

Qy        241 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 640

Qy        301 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        641 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCCCGCAACGTGGGT 700

Qy        361 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        701 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTA 760

Qy        421 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        761 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 820

Qy        481 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        821 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 880

Qy        541 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        881 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 940

Qy        601 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 660
              |||||||||| |||||||||||||||||| |||||||||||| ||||||||||||||||
Db        941 TACATGGTGACCAACGACTGCACCAATGATAGCATTACCTGGCAACTCCAGGCTGCTGTC 1000

Qy        661 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 720
              |||||||||||||||||||||||||||||||||||||||||||||| |||||| ||||||||||
Db       1001 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATACATCTCGGTGCTGGATA 1060

Qy        721 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 780
              |||||||||||||||||||||||||||||| |||||||||||||||||||||| ||||||
Db       1061 CCGGTCTCACCGAATGTGGCCGTGCAGCAGCCCGGCGCCCTCACGCAGGGCTTACGGACG 1120

Qy        781 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 840
              ||||| |||||||||||||||||||||||||||||||||| || || ||||||||||||||
Db       1121 CACATTGACATGGTTGTGATGTCCGCCACGCTCTGCTCCGCTCTTTACGTGGGGGACCTC 1180

Qy        841 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 900
              |||||||||||||||||| |||||||| |||||||||||||||| |||||||||||||||
Db       1181 TGCGGTGGGGTGATGCTTGCAGCCCAGATGTTCATTGTCTCGCCACAGCACCACTGGTTT 1240

Qy        901 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 960
              || |||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db       1241 GTGCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCGTGG 1300

Qy        961 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1020
              |||||||||||||||||||||||||||||||||||||||| |||||||||||||| |||
Db       1301 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCCTGGCGTACGCGATGCGCGTC 1360

Qy       1021 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1080
              |||||||||| |||||||||| |||||||||||||||||||||||||||||||| |||
Db       1361 CCCGAGGTCATCATAGACATCATTGGCGGGGCTCATTGGGGCGTCATGTTCGGCTTAGCC 1420

Qy       1081 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1140
              |||||||||||||||||||||||| |||||||||||||| ||| || ||||||||||||
Db       1421 TACTTCTCTATGCAGGGAGCGTGGGCAAAAGTCGTTGTCATTCTTTTGCTGGCCGCCGGG 1480

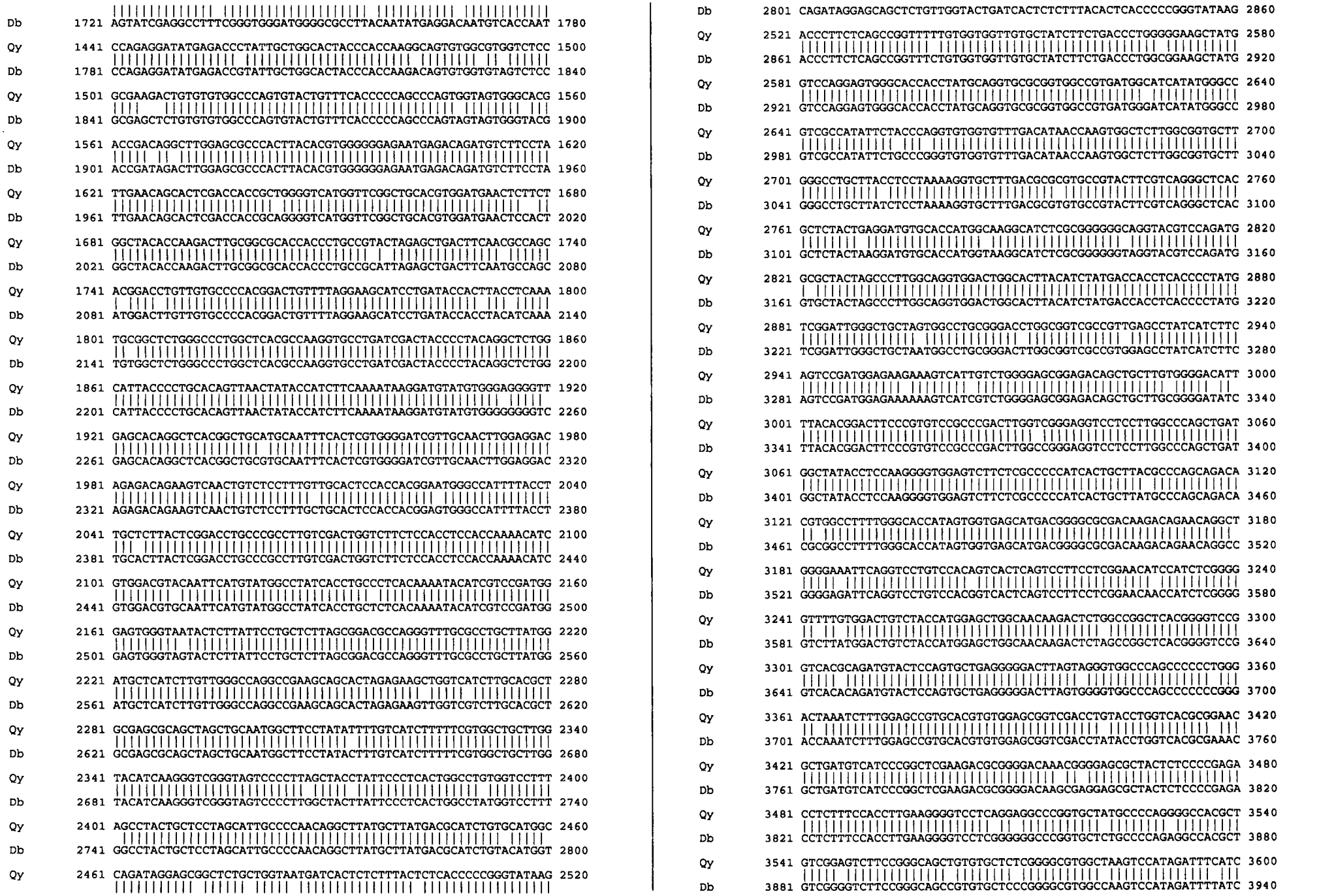
Qy       1141 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1200
              |||||||||| |||||||| |||||||||||| ||||||| | | || ||  |||||||
Db       1481 GTGGACGCGCAAACCCATACCGTTGGGGGTTCTACCGCGCATAACGCCAGGACCCTCACC 1540

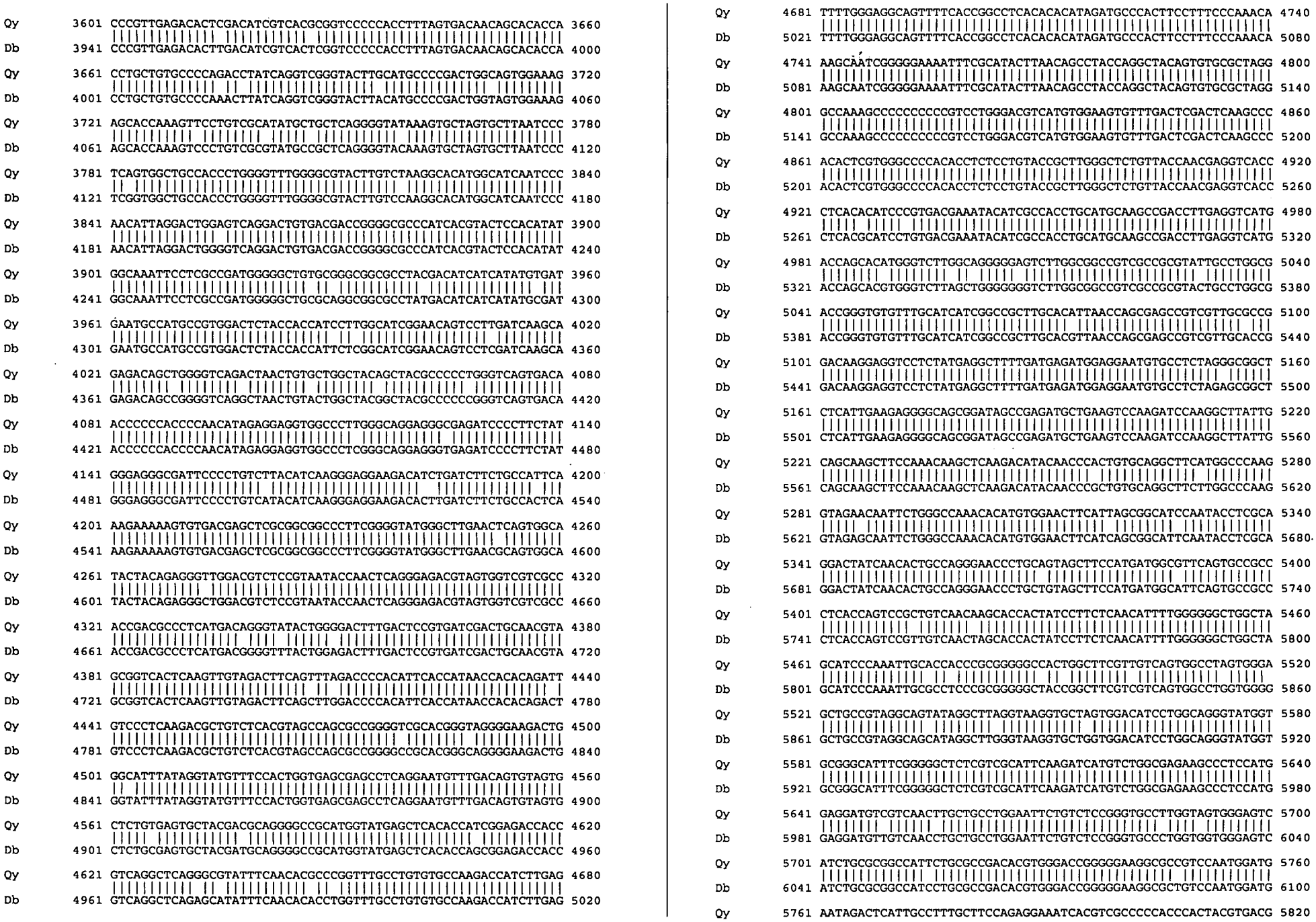
Qy       1201 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1260
              || | ||   | | ||  |||||||||||||||||| | ||||||||||||| |||
Db       1541 GGCATGTTCTCCCTTGGTGCCAGGCAGAAAATCCAGCTCATCAACACCAATGGCAGTTGG 1600

Qy       1261 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1320
              |||||||||||||||||||||||||||||||||| ||||||||||||||  |||||||
Db       1601 CACATCAACCGCACCGCCCTGAACTGCAATGACTCTTTGCACACCGGCTTCCTCGCGTCA 1660

Qy       1321 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1380
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1661 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1720

Qy       1381 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1440
```


Db 1721 AGTATCGAGGCCTTTCGGGTGGGATGGGGCGCCTTACAATATGAGGACAATGTCACCAAT 1780
Qy 1441 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1500
Db 1781 CCAGAGGATATGAGACCGTATTGCTGGCACTACCCACCAAGACAGTGTGGTGTAGTCTCC 1840
Qy 1501 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1560
Db 1841 GCGAGCTCTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTAGTAGTGGGTACG 1900
Qy 1561 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1620
Db 1901 ACCGATAGACTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1960
Qy 1621 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 1680
Db 1961 TTGAACAGCACTCGACCACCGCAGGGGTCATGGTTCGGCTGCACGTGGATGAACTCCACT 2020
Qy 1681 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 1740
Db 2021 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGCATTAGAGCTGACTTCAATGCCAGC 2080
Qy 1741 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 1800
Db 2081 ATGGACTTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACCTACATCAAA 2140
Qy 1801 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 1860
Db 2141 TGTGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 2200
Qy 1861 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 1920
Db 2201 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGGGGGGTC 2260
Qy 1921 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 1980
Db 2261 GAGCACAGGCTCACGGCTGCGTGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 2320
Qy 1981 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2040
Db 2321 AGAGACAGAAGTCAACTGTCTCCTTTGCTGCACTCCACCACGGAGTGGGCCATTTTACCT 2380
Qy 2041 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2100
Db 2381 TGCACTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2440
Qy 2101 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2160
Db 2441 GTGGACGTGCAATTCATGTATGGCCTATCACCTGCTCTCACAAAATACATCGTCCGATGG 2500
Qy 2161 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2220
Db 2501 GAGTGGGTAGTACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2560
Qy 2221 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2280
Db 2561 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGTTGGTCGTCTTGCACGCT 2620
Qy 2281 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATATTTTGTCATCTTTTTCGTGGCTGCTTGG 2340
Db 2621 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATACTTTGTCATCTTTTTCGTGGCTGCTTGG 2680
Qy 2341 TACATCAAGGGTCGGGTAGTCCCCTTAGCTACCTATTCCCTCACTGGCCTGTGGTCCTTT 2400
Db 2681 TACATCAAGGGTCGGGTAGTCCCCTTGGCTACTTATTCCCTCACTGGCCTATGGTCCTTT 2740
Qy 2401 AGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTGCATGGC 2460
Db 2741 GGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTACATGGT 2800
Qy 2461 CAGATAGGAGCGGCTCTGCTGGTAATGATCACTCTCTTTACTCTCACCCCCGGGTATAAG 2520
Db 2801 CAGATAGGAGCAGCTCTGTTGGTACTGATCACTCTCTTTACACTCACCCCCGGGTATAAG 2860
Qy 2521 ACCCTTCTCAGCCGGTTTTTGTGGTGGTTGTGCTATCTTCTGACCCTGGGGGAAGCTATG 2580
Db 2861 ACCCTTCTCAGCCGGTTTCTGTGGTGGTTGTGCTATCTTCTGACCCTGGCGGAAGCTATG 2920
Qy 2581 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGCATCATATGGGCC 2640
Db 2921 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGGATCATATGGGCC 2980
Qy 2641 GTCGCCATATTCTACCCAGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 2700
Db 2981 GTCGCCATATTCTGCCCGGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 3040
Qy 2701 GGGCCTGCTTACCTCCTAAAAGGTGCTTTGACGCGCGTGCCGTACTTCGTCAGGGCTCAC 2760
Db 3041 GGGCCTGCTTATCTCCTAAAAGGTGCTTTGACGCGTGTGCCGTACTTCGTCAGGGCTCAC 3100
Qy 2761 GCTCTACTGAGGATGTGCACCATGGCAAGGCATCTCGCGGGGGGCAGGTACGTCCAGATG 2820
Db 3101 GCTCTACTAAGGATGTGCACCATGGTAAGGCATCTCGCGGGGGGTAGGTACGTCCAGATG 3160
Qy 2821 GCGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 2880
Db 3161 GTGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 3220
Qy 2881 TCGGATTGGGCTGCTAGTGGCCTGCGGGACCTGGCGGTCGCCGTTGAGCCTATCATCTTC 2940
Db 3221 TCGGATTGGGCTGCTAATGGCCTGCGGGACTTGGCGGTCGCCGTGGAGCCTATCATCTTC 3280
Qy 2941 AGTCCGATGGAGAAGAAAGTCATTGTCTGGGGAGCGGAGACAGCTGCTTGTGGGGACATT 3000
Db 3281 AGTCCGATGGAGAAAAAAGTCATCGTCTGGGGAGCGGAGACAGCTGCTTGCGGGGATATC 3340
Qy 3001 TTACACGGACTTCCCGTGTCCGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3060
Db 3341 TTACACGGACTTCCCGTGTCCGCCCGACTTGGCCGGGAGGTCCTCCTTGGCCCAGCTGAT 3400
Qy 3061 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3120
Db 3401 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTATGCCCAGCAGACA 3460
Qy 3121 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCT 3180
Db 3461 CGCGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCC 3520
Qy 3181 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3240
Db 3521 GGGGAGATTCAGGTCCTGTCCACGGTCACTCAGTCCTTCCTCGGAACAACCATCTCGGGG 3580
Qy 3241 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3300
Db 3581 GTCTTATGGACTGTCTACCATGGAGCTGGCAACAAGACTCTAGCCGGCTCACGGGGTCCG 3640
Qy 3301 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGG 3360
Db 3641 GTCACACAGATGTACTCCAGTGCTGAGGGGGACTTAGTGGGGTGGCCCAGCCCCCCCGGG 3700
Qy 3361 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3420
Db 3701 ACCAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTATACCTGGTCACGCGAAAC 3760
Qy 3421 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3480
Db 3761 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAGCGAGGAGCGCTACTCTCCCCGAGA 3820
Qy 3481 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3540
Db 3821 CCTCTTTCCACCTTGAAGGGGTCCTCGGGGGGCCCGGTGCTCTGCCCCAGAGGCCACGCT 3880
Qy 3541 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3600
Db 3881 GTCGGGGTCTTCCGGGCAGCCGTGTGCTCCCGGGGCGTGGCCAAGTCCATAGATTTTATC 3940




Db 6101 AATAGGCTCATTGCCTTTGCTTCCAGAGGAAACCACGTCGCCCCCACCCACTACGTGACG 6160
Qy 5821 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 5880
Db 6161 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 6220
Qy 5881 CTCAGAAGACTCCACAACTGGATTACTGAGGACTGCCCCATCCCATGCGGCGGCTCGTGG 5940
Db 6221 CTCAGGAGACTCCACAACTGGATTACTGAAGACTGCCCCATCCCATGCAGCGGCTCGTGG 6280
Qy 5941 CTCCGCGATGTGTGGGACTGGGTTTGCACCATCCTAACAGACTTTAAAAATTGGCTGACC 6000
Db 6281 CTCCGCGATGTGTGGGATTGGGTTTGCACCATCCTAACAGACTTTAAAAACTGGCTGACC 6340
Qy 6001 TCCAAATTATTCCCAAAGATGCCCGGCCTCCCCTTTGTCTCCTGTCAAAAGGGGTACAAG 6060
Db 6341 TCCAAATTGTTCCCAAAGATGCCTGGTCTCCCCTTTATCTCTTGTCAAAAGGGGTACAAG 6400
Qy 6061 GGCGTGTGGGCCGGCACTGGCATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6120
Db 6401 GGCGTGTGGGCTGGCACTGGTATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6460
Qy 6121 GGCAATGTCCGCTTGGGCTCCATGAGAATCACGGGGCCTAAGACCTGCATGAATATCTGG 6180
Db 6461 GGCAATGTCCGCCTGGGCTCCATGAGAATTACGGGGCCCAAAACCTGCATGAATATCTGG 6520
Qy 6181 CAGGGGACCTTTCCTATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCGCCA 6240
Db 6521 CAGGGGACCTTTCCCATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCACCA 6580
Qy 6241 AACTTTAAGGTCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6300
Db 6581 AACTTTAAGATCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6640
Qy 6301 GGGTCATACCACTACATAACAGGACTCACCACTGATAACTTGAAAGTCCCCTGCCAACTA 6360
Db 6641 GGGTCATACCACTACATAACAGGACTTACCACTGATAACTTGAAAGTTCCTTGCCAACTA 6700
Qy 6361 CCCTCTCCCGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCACA 6420
Db 6701 CCTTCTCCAGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCATA 6760
Qy 6421 CCGAAGCCGTTTTTCCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6480
Db 6761 CCGAAGCCGTTTTTTCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6820
Qy 6481 GGGTCCCAGCTTCCTTGCGACCCTGAACCCGACACAGACGTATTGATGTCCATGCTAACA 6540
Db 6821 GGGTCTCAGCTCCCTTGCGATCCTGAACCTGACACAGACGTATTGACGTCCATGCTAACA 6880
Qy 6541 GATCCATCTCATATCACGGCGGAGACTGCAGCGCGGCGTTTAGCGCGGGGGTCACCCCCA 6600
Db 6881 GACCCATCCCATATCACGGCGGAGACTGCAGCGCGGCGTTTGGCACGGGGGTCACCCCCG 6940
Qy 6601 TCCGAGGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 6660
Db 6941 TCCGAGGCAAGCTCCTCAGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 7000
Qy 6661 ACCCACGGCAAAGCCTATGATGTGGACATGGTGGATGCTAACCTGTTCATGGGGGGCGAT 6720
Db 7001 ACCCACGGCAAGGCCTATGATGTGGACATGGTGGATGCCAACCTGTTCATGGGGGGCGAT 7060
Qy 6721 GTGACTCGGATAGAGTCTGGGTCCAAAGTGGTCGTTCTGGACTCTCTCGACCCAATGGTC 6780
Db 7061 GTGACCCGGATAGAGTCTGAGTCCAAAGTGGTCGTTCTGGACTCTCTCGACCCAATGGTC 7120
Qy 6781 GAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCAGAATACATGCTCCCCAAGAAGAGG 6840
Db 7121 GAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCGGAATATATGCTCCCCAAGAAGAGA 7180
Qy 6841 TTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCGCTTGTGGAATCG 6900
Db 7181 TTCCCACCAGCCTTACCGGCTTGGGCACGGCCTGATTACAACCCACCGCTTGTGGAATCG 7240
Qy 6901 TGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTCCCTCCTCCTAGG 6960
Db 7241 TGGAAGAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGCGCTCTCCCCCCCCCTAAG 7300
Qy 6961 AAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGTGAGGACTCCATA 7020
Db 7301 AAAACCCCGACGCCTCCCCCAAGGAGACGCCGGACAGTGGGTCTGAGTGAGAGCTCCATA 7360
Qy 7021 GGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCCCCAAGCGGCGAT 7080
Db 7361 GCAGATGCCCTACAACAGCTGGCCATCAAGTCCTTTGGCCAGCCCCCCCCAAGCGGCGAT 7420
Qy 7081 TCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACGCCTCCTGATGAG 7140
Db 7421 TCAGGCCTTTCCACGGGGGCGGACGCAGCCGATTCCGGCAGTCGGACGCCCCCCGATGAG 7480
Qy 7141 TTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCTCGAGGGGGAGCTTGGA 7200
Db 7481 TTGGCCCTTTCGGAGACAGGTTCCATCTCCTCCATGCCCCCTCTCGAGGGGGAGCCTGGA 7540
Qy 7201 GATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCCAGGGGGGGGTGGCA 7260
Db 7541 GATCCAGACTTGGAGCCTGAGCAGGTAGAGCTTCAACCTCCCCCCCCAGGGGGGGGTGGTA 7600
Qy 7261 GCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGACGACTCCGTCGTG 7320
Db 7601 ACCCCCGGCTCAGGCTCGGGGTCTTGGTCTACTTGCTCCGAGGAGGACGACTCCGTCGTG 7660
Qy 7321 TGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGTAGTCCCGAAGAG 7380
Db 7661 TGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGTAGCCCCGAAGAG 7720
Qy 7381 GAGAAGTTACCGATTAACCCCTTGAGCAACTCCCTGTTGCGATATCACAACAAGGTGTAC 7440
Db 7721 GAAAAGTTGCCAATTGGCCCCTTGAGCAACTCCCTGTTGCGATATCACAACAAGGTGTAC 7780
Qy 7441 TGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTTGATAGGATGCAA 7500
Db 7781 TGTACCACATCAAAGAGCGCCTCATTAAGGGCTAAAAAGGTAACTTTTGATAGGATGCAA 7840
Qy 7501 GTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCGGCCTCCAAGGTC 7560
Db 7841 GCGCTCGACGCTCATTATGACTCAGTCTTGAAGGACATTAAGCTAGCGGCCTCCAAGGTC 7900
Qy 7561 ACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCCCATTCTGCAAGA 7620
Db 7901 ACCGCAAGGCTTCTCACTTTAGAGGAGGCCTGCCAGTTAACTCCACCCCACTCTGCAAGA 7960
Qy 7621 TCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGGGCCGTTAACCAC 7680
Db 7961 TCCAAGTATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGAGCCGTTAACCAC 8020
Qy 7681 ATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATTCCCACAACCATT 7740
Db 8021 ATCAAGTCCGTGTGGAAGGACCTCCTGGAAGACACACAAACACCAATTCCTACAACCATC 8080
Qy 7741 ATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAGAAAGCAGCTCGC 7800
Db 8081 ATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGTAAGAAAGCAGCTCGC 8140
Qy 7801 CTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCCCTTTATGACATT 7860
Db 8141 CTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAAATGGCCCTTTATGATATC 8200
Qy 7861 ACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAGTATTCCCCCGCT 7920
Db 8201 ACACAAAAGCTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAGTACTCCCCCGCT 8260
Qy 7921 CAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCTATGGGTTTTTCG 7980
Db 8261 CAGCGGGTGGAGTTTCTCTTGAAGGCATGGGCGGAAAAGAAAGACCCTATGGGTTTTTCG 8320

```
Qy       7981 TATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGGACTGAGGAGTCC 8040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8321 TATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGGACTGAGGAGTCC 8380

Qy       8041 ATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACACTCGCTAACTGAG 8100
             |||||||||||  || ||||||||||||||||||||||||||||||||  || ||||||
Db       8381 ATATATCGGGCTTGTTCCTTGCCCGAGGAGGCCCACACTGCCATACACTCACTGACTGAG 8440

Qy       8101 AGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGCGGGTACAGGCGT 8160
             |||||||||||||||||||  ||||||||||||||||||||  ||||||||||||||||||
Db       8441 AGACTTTACGTGGGAGGGCCCATGTTCAACAGCAAGGGCCAGACCTGCGGGTACAGGCGT 8500

Qy       8161 TGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACATGCTACGTGAAA 8220
             |||||||||||||||||||  |||||||||||||||||||||||||||||||||  ||||||
Db       8501 TGCCGCGCCAGCGGGGTGCTTACCACTAGCATGGGGAACACCATCACATGCTATGTGAAA 8560

Qy       8221 GCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTGGTATGCGGCGAT 8280
             ||||||||||| ||||| ||||||||||||| ||||||||||||||||||||||||||||||
Db       8561 GCCTTAGCGGCCTGTAAGGCTGCAGGGATAATTGCGCCCACAATGCTGGTATGCGGCGAT 8620

Qy       8281 GACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGGAACCTGAGAGCC 8340
             |||||||||||||||||||  |||||||||||||||||||||||||||||||||||||||||
Db       8621 GACTTGGTTGTCATCTCAGAGAGCCAGGGGACCGAGGAGGACGAGCGGAACCTGAGAGCC 8680

Qy       8341 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCCAGACCGGAGTAT 8400
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||  |||
Db       8681 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCCAGACCGGAATAT 8740

Qy       8401 GATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTGGGCCCACAAGGC 8460
             || ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db       8741 GACCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGTTGGGCCCACAAGGC 8800

Qy       8461 CGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGGGCTGCCTGGGAA 8520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8801 CGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGGGCTGCCTGGGAA 8860

Qy       8521 ACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAGTACGCCCCGACC 8580
             |||||||||||||||||||||||||||||||| |||||||||||||||||||||||| |||
Db       8861 ACAGTTAGACACTCCCCTGTCAATTCATGGCTAGGAAACATCATCCAGTACGCCCCAACC 8920

Qy       8581 ATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATGGCTCAAGACACG 8640
             |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||| ||
Db       8921 ATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATGGCCCAAGATACT 8980

Qy       8641 CTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTGAGTCCCTTGGAC 8700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8981 CTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTGAGTCCCTTGGAC 9040

Qy       8701 CTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTGCACACATACACT 8760
             |||||||| |||||||||||||||| |||||||||||| ||||||||||||||||||||||||
Db       9041 CTCCCAGCCATAATTGAAAGGTTACACGGGCTTGACGCTTTCTCTCTGCACACATACACT 9100

Qy       8761 CCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGA 8820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9101 CCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGA 9160

Qy       8821 GCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGTGGGGGGAGAGCG 8880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9161 GCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGTGGGGGGAGAGCG 9220

Qy       8881 GCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 8940
             |||||||||||| |||||||||||| ||||||||||||||||||||||||||||||||||||
Db       9221 GCCGTTTGCGGCCGATATCTCTTCAACTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 9280

Qy       8941 TTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGCGCCGGCGGGGGC 9000
             |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db       9281 TTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACTGTCGGCGCCGGCGGGGGC 9340

Qy       9001 GACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTTGGCCTACTCCTA 9060
             |||||||||||||||||||||||||||||||||||||||||||||| ||| ||||||||||||
Db       9341 GACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTACTCCTTGGCCTACTCCTA 9400

Qy       9061 CTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCGG 9099
             |||||||||||||||||||||||||||||||||||||||
Db       9401 CTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCGG 9439


RESULT 8
I12861
LOCUS       I12861                 9589 bp    DNA     linear   PAT 26-JUL-1995
DEFINITION  Sequence 2 from patent US 5428145.
ACCESSION   I12861
VERSION     I12861.1  GI:910242
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 9589)
  AUTHORS   Okamoto,H. and Nakamura,T.
  TITLE     Non-A, non-B, hepatitis virus genome, polynucleotides,
            polypeptides, antigen, antibody and detection systems
  JOURNAL   Patent: US 5428145-A 2 27-JUN-1995;
FEATURES             Location/Qualifiers
     source          1. .9589
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             93.4%;  Score 8494.2;  DB 6;  Length 9589;
  Best Local Similarity   95.8%;  Pred. No. 0;
  Matches 8721;  Conservative    0;  Mismatches  378;  Indels    0;  Gaps    0;

Qy          1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 400

Qy         61 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 460

Qy        121 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 520

Qy        181 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 240
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        521 AGGCGCCAGCCCATCCCTAAGGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 580

Qy        241 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 640

Qy        301 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 360
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        641 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCCCGCAACGTGGGT 700

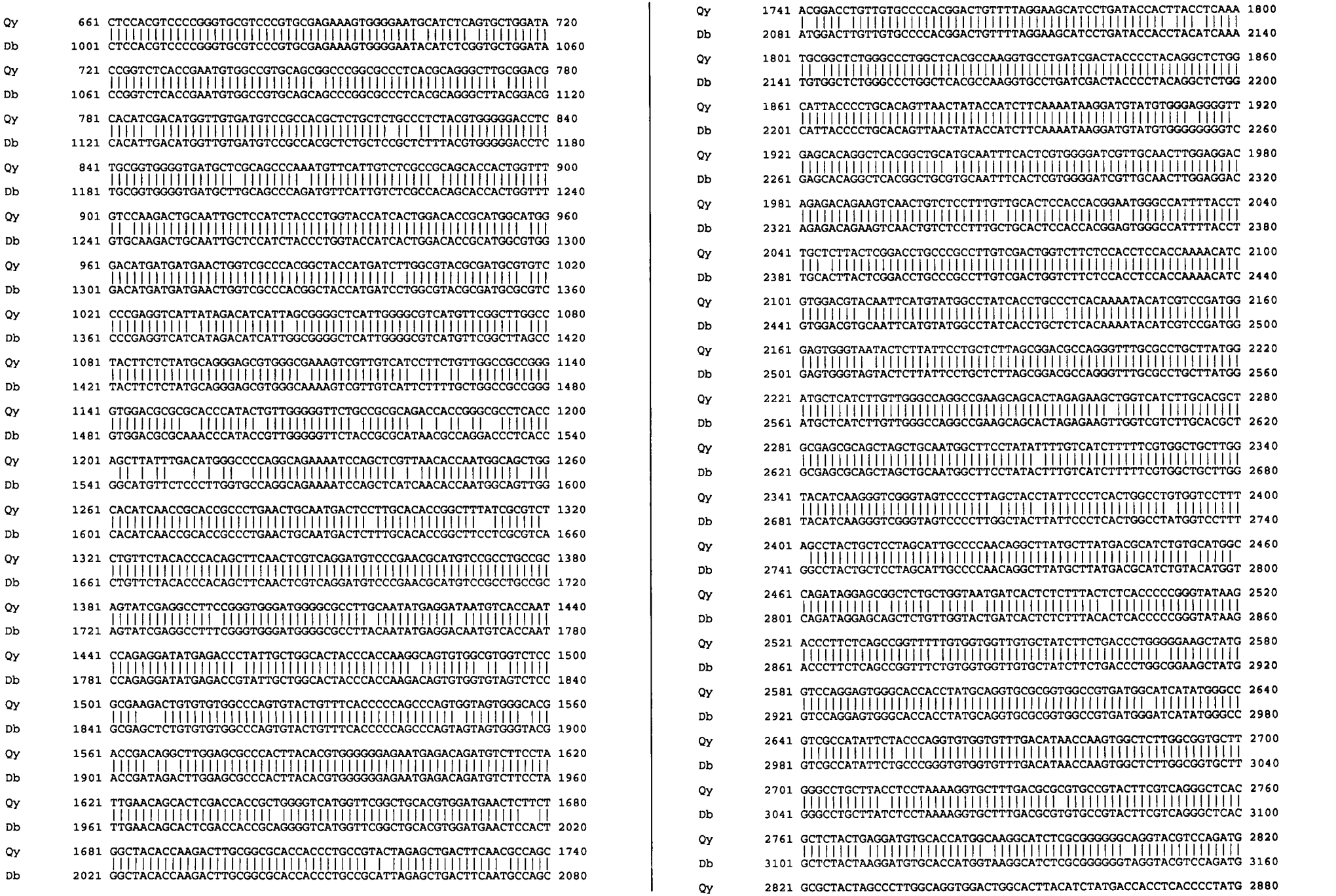
Qy        361 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        701 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTA 760

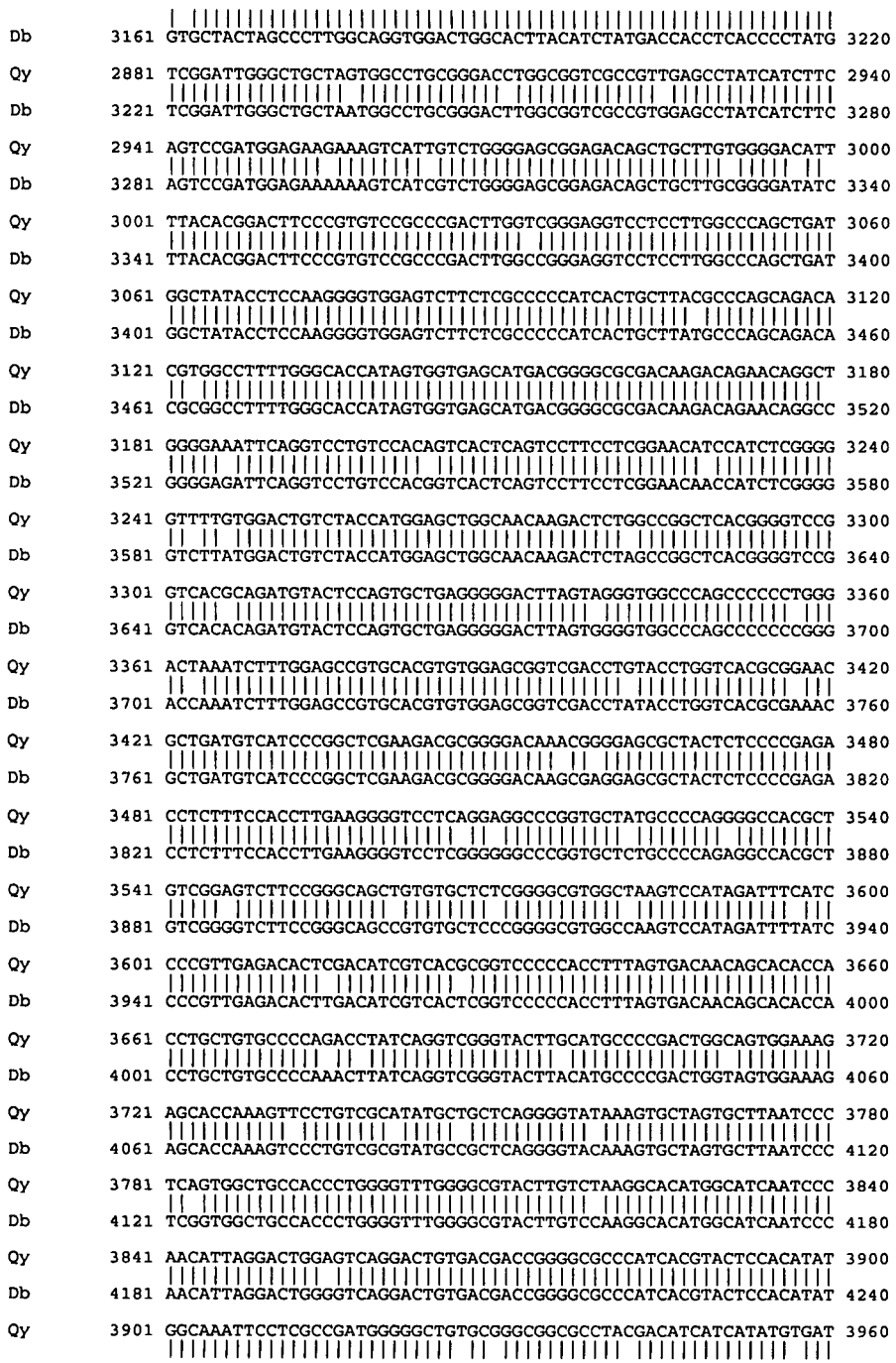
Qy        421 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        761 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 820

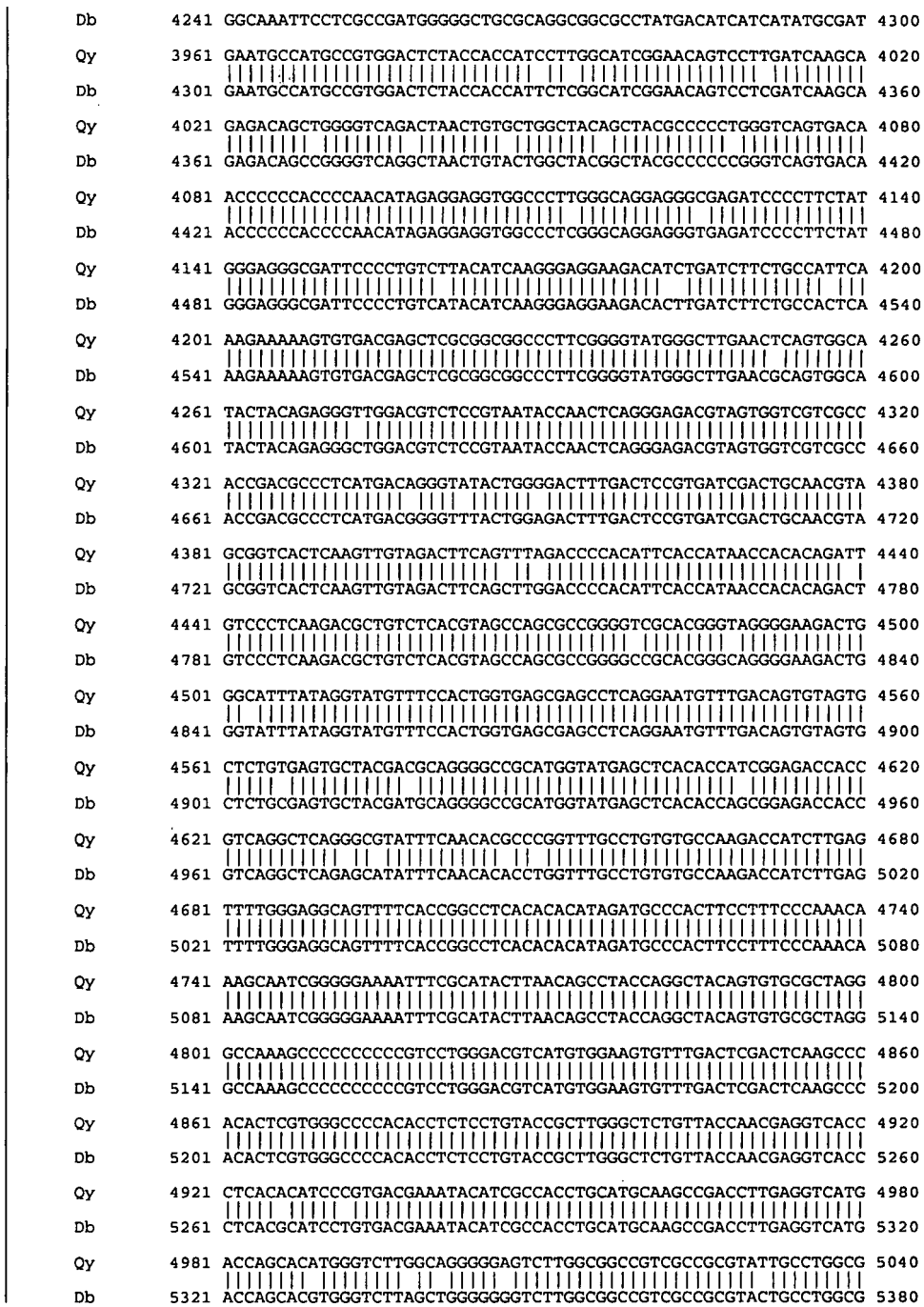
Qy        481 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        821 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 880

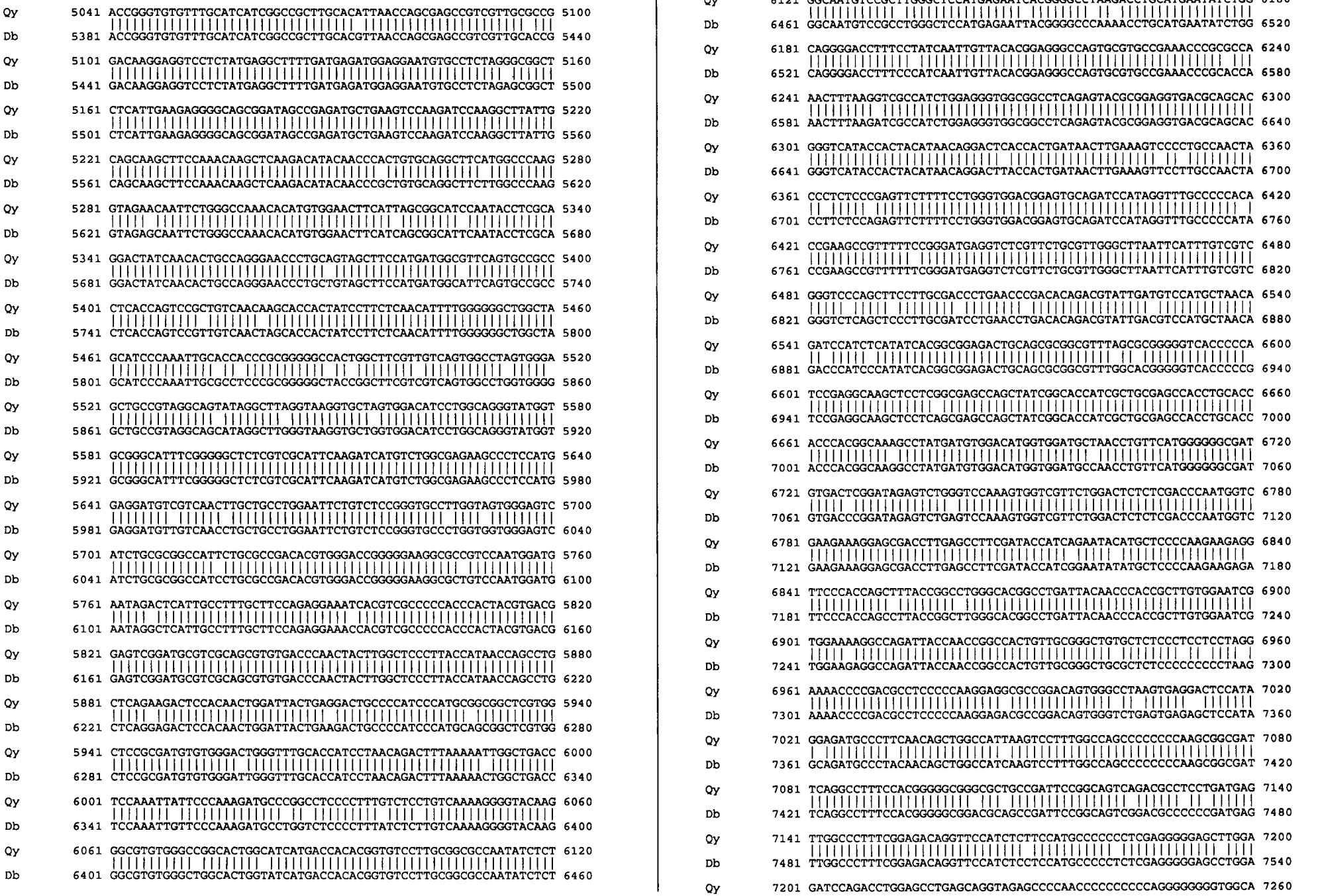
Qy        541 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        881 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 940

Qy        601 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 660
              |||||||||||| |||||||||||||||| |||||||||||| ||||||||||||||||
Db        941 TACATGGTGACCAACGACTGCACCAATGATAGCATTACCTGGCAACTCCAGGCTGCTGTC 1000
```




Db 3161 GTGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 3220
Qy 2881 TCGGATTGGGCTGCTAGTGGCCTGCGGGACCTGGCGGTCGCCGTTGAGCCTATCATCTTC 2940
Db 3221 TCGGATTGGGCTGCTAATGGCCTGCGGGACTTGGCGGTCGCCGTGGAGCCTATCATCTTC 3280
Qy 2941 AGTCCGATGGAGAAGAAAGTCATTGTCTGGGGAGCGGAGACAGCTGCTTGTGGGGACATT 3000
Db 3281 AGTCCGATGGAGAAAAAAGTCATCGTCTGGGGAGCGGAGACAGCTGCTTGCGGGGATATC 3340
Qy 3001 TTACACGGACTTCCCGTGTCCGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3060
Db 3341 TTACACGGACTTCCCGTGTCCGCCCGACTTGGCCGGGAGGTCCTCCTTGGCCCAGCTGAT 3400
Qy 3061 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3120
Db 3401 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTATGCCCAGCAGACA 3460
Qy 3121 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCT 3180
Db 3461 CGCGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCC 3520
Qy 3181 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3240
Db 3521 GGGGAGATTCAGGTCCTGTCCACGGTCACTCAGTCCTTCCTCGGAACAACCATCTCGGGG 3580
Qy 3241 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3300
Db 3581 GTCTTATGGACTGTCTACCATGGAGCTGGCAACAAGACTCTAGCCGGCTCACGGGGTCCG 3640
Qy 3301 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGG 3360
Db 3641 GTCACACAGATGTACTCCAGTGCTGAGGGGGACTTAGTGGGGTGGCCCAGCCCCCCCGGG 3700
Qy 3361 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3420
Db 3701 ACCAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTATACCTGGTCACGCGAAAC 3760
Qy 3421 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3480
Db 3761 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAGCGAGGAGCGCTACTCTCCCCGAGA 3820
Qy 3481 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3540
Db 3821 CCTCTTTCCACCTTGAAGGGGTCCTCGGGGGGCCCGGTGCTCTGCCCCAGAGGCCACGCT 3880
Qy 3541 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3600
Db 3881 GTCGGGGTCTTCCGGGCAGCCGTGTGCTCCCGGGGCGTGGCCAAGTCCATAGATTTTATC 3940
Qy 3601 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 3660
Db 3941 CCCGTTGAGACACTTGACATCGTCACTCGGTCCCCCACCTTTAGTGACAACAGCACACCA 4000
Qy 3661 CCTGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 3720
Db 4001 CCTGCTGTGCCCCAAACTTATCAGGTCGGGTACTTACATGCCCCGACTGGTAGTGGAAAG 4060
Qy 3721 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 3780
Db 4061 AGCACCAAAGTCCCTGTCGCGTATGCCGCTCAGGGGTACAAAGTGCTAGTGCTTAATCCC 4120
Qy 3781 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 3840
Db 4121 TCGGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCCAAGGCACATGGCATCAATCCC 4180
Qy 3841 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 3900
Db 4181 AACATTAGGACTGGGGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 4240
Qy 3901 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 3960


Db 4241 GGCAAATTCCTCGCCGATGGGGGCTGCGCAGGCGGCGCCTATGACATCATCATATGCGAT 4300
Qy 3961 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4020
Db 4301 GAATGCCATGCCGTGGACTCTACCACCATTCTCGGCATCGGAACAGTCCTCGATCAAGCA 4360
Qy 4021 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4080
Db 4361 GAGACAGCCGGGGTCAGGCTAACTGTACTGGCTACGGCTACGCCCCCCGGGTCAGTGACA 4420
Qy 4081 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4140
Db 4421 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTCGGGCAGGAGGGTGAGATCCCCTTCTAT 4480
Qy 4141 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4200
Db 4481 GGGAGGGCGATTCCCCTGTCATACATCAAGGGAGGAAGACACTTGATCTTCTGCCACTCA 4540
Qy 4201 AAGAAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4260
Db 4541 AAGAAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACGCAGTGGCA 4600
Qy 4261 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4320
Db 4601 TACTACAGAGGGCTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4660
Qy 4321 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4380
Db 4661 ACCGACGCCCTCATGACGGGGTTTACTGGAGACTTTGACTCCGTGATCGACTGCAACGTA 4720
Qy 4381 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4440
Db 4721 GCGGTCACTCAAGTTGTAGACTTCAGCTTGGACCCCACATTCACCATAACCACACAGACT 4780
Qy 4441 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4500
Db 4781 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGCCGCACGGGCAGGGGAAGACTG 4840
Qy 4501 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4560
Db 4841 GGTATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4900
Qy 4561 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4620
Db 4901 CTCTGCGAGTGCTACGATGCAGGGGCCGCATGGTATGAGCTCACACCAGCGGAGACCACC 4960
Qy 4621 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 4680
Db 4961 GTCAGGCTCAGAGCATATTTCAACACACCTGGTTTGCCTGTGTGCCAAGACCATCTTGAG 5020
Qy 4681 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 4740
Db 5021 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 5080
Qy 4741 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 4800
Db 5081 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 5140
Qy 4801 GCCAAAGCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 4860
Db 5141 GCCAAAGCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 5200
Qy 4861 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 4920
Db 5201 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 5260
Qy 4921 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 4980
Db 5261 CTCACGCATCCTGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 5320
Qy 4981 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5040
Db 5321 ACCAGCACGTGGGTCTTAGCTGGGGGGGTCTTGGCGGCCGTCGCCGCGTACTGCCTGGCG 5380


Qy 5041 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5100
Db 5381 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACGTTAACCAGCGAGCCGTCGTTGCACCG 5440
Qy 5101 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5160
Db 5441 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGAGCGGCT 5500
Qy 5161 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5220
Db 5501 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5560
Qy 5221 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5280
Db 5561 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCGCTGTGCAGGCTTCTTGGCCCAAG 5620
Qy 5281 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5340
Db 5621 GTAGAGCAATTCTGGGCCAAACACATGTGGAACTTCATCAGCGGCATTCAATACCTCGCA 5680
Qy 5341 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5400
Db 5681 GGACTATCAACACTGCCAGGGAACCCTGCTGTAGCTTCCATGATGGCATTCAGTGCCGCC 5740
Qy 5401 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5460
Db 5741 CTCACCAGTCCGTTGTCAACTAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5800
Qy 5461 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5520
Db 5801 GCATCCCAAATTGCGCCTCCCGCGGGGGCTACCGGCTTCGTCGTCAGTGGCCTGGTGGGG 5860
Qy 5521 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5580
Db 5861 GCTGCCGTAGGCAGCATAGGCTTGGGTAAGGTGCTGGTGGACATCCTGGCAGGGTATGGT 5920
Qy 5581 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5640
Db 5921 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5980
Qy 5641 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 5700
Db 5981 GAGGATGTTGTCAACCTGCTGCCTGGAATTCTGTCTCCGGGTGCCCTGGTGGTGGGAGTC 6040
Qy 5701 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 5760
Db 6041 ATCTGCGCGGCCATCCTGCGCCGACACGTGGGACCGGGGGAAGGCGCTGTCCAATGGATG 6100
Qy 5761 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 5820
Db 6101 AATAGGCTCATTGCCTTTGCTTCCAGAGGAAACCACGTCGCCCCCACCCACTACGTGACG 6160
Qy 5821 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 5880
Db 6161 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 6220
Qy 5881 CTCAGAAGACTCCACAACTGGATTACTGAGGACTGCCCCATCCCATGCGGCGGCTCGTGG 5940
Db 6221 CTCAGGAGACTCCACAACTGGATTACTGAAGACTGCCCCATCCCATGCAGCGGCTCGTGG 6280
Qy 5941 CTCCGCGATGTGTGGGACTGGGTTTGCACCATCCTAACAGACTTTAAAAATTGGCTGACC 6000
Db 6281 CTCCGCGATGTGTGGGATTGGGTTTGCACCATCCTAACAGACTTTAAAAACTGGCTGACC 6340
Qy 6001 TCCAAATTATTCCCAAAGATGCCCGGCCTCCCCTTTGTCTCCTGTCAAAAGGGGTACAAG 6060
Db 6341 TCCAAATTGTTCCCAAAGATGCCTGGTCTCCCCTTTATCTCTTGTCAAAAGGGGTACAAG 6400
Qy 6061 GGCGTGTGGGCCGGCACTGGCATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6120
Db 6401 GGCGTGTGGGCTGGCACTGGTATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6460
Qy 6121 GGCAATGTCCGCTTGGGCTCCATGAGAATCACGGGGCCTAAGACCTGCATGAATATCTGG 6180
Db 6461 GGCAATGTCCGCCTGGGCTCCATGAGAATTACGGGGCCCAAAACCTGCATGAATATCTGG 6520
Qy 6181 CAGGGGACCTTTCCTATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCGCCA 6240
Db 6521 CAGGGGACCTTTCCCATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCACCA 6580
Qy 6241 AACTTTAAGGTCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6300
Db 6581 AACTTTAAGATCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6640
Qy 6301 GGGTCATACCACTACATAACAGGACTCACCACTGATAACTTGAAAGTCCCCTGCCAACTA 6360
Db 6641 GGGTCATACCACTACATAACAGGACTTACCACTGATAACTTGAAAGTTCCTTGCCAACTA 6700
Qy 6361 CCCTCTCCCGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCACA 6420
Db 6701 CCTTCTCCAGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCATA 6760
Qy 6421 CCGAAGCCGTTTTTCCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6480
Db 6761 CCGAAGCCGTTTTTTCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6820
Qy 6481 GGGTCCCAGCTTCCTTGCGACCCTGAACCCGACACAGACGTATTGATGTCCATGCTAACA 6540
Db 6821 GGGTCTCAGCTCCCTTGCGATCCTGAACCTGACACAGACGTATTGACGTCCATGCTAACA 6880
Qy 6541 GATCCATCTCATATCACGGCGGAGACTGCAGCGCGGCGTTTAGCGCGGGGGTCACCCCCA 6600
Db 6881 GACCCATCCCATATCACGGCGGAGACTGCAGCGCGGCGTTTGGCACGGGGGTCACCCCCG 6940
Qy 6601 TCCGAGGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 6660
Db 6941 TCCGAGGCAAGCTCCTCAGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 7000
Qy 6661 ACCCACGGCAAAGCCTATGATGTGGACATGGTGGATGCTAACCTGTTCATGGGGGGCGAT 6720
Db 7001 ACCCACGGCAAGGCCTATGATGTGGACATGGTGGATGCCAACCTGTTCATGGGGGGCGAT 7060
Qy 6721 GTGACTCGGATAGAGTCTGGGTCCAAAGTGGTCGTTCTGGACTCTCTCGACCCAATGGTC 6780
Db 7061 GTGACCCGGATAGAGTCTGAGTCCAAAGTGGTCGTTCTGGACTCTCTCGACCCAATGGTC 7120
Qy 6781 GAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCAGAATACATGCTCCCCAAGAAGAGG 6840
Db 7121 GAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCGGAATATATGCTCCCCAAGAAGAGA 7180
Qy 6841 TTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCGCTTGTGGAATCG 6900
Db 7181 TTCCCACCAGCCTTACCGGCTTGGGCACGGCCTGATTACAACCCACCGCTTGTGGAATCG 7240
Qy 6901 TGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTCCCTCCTCCTAGG 6960
Db 7241 TGGAAGAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGCGCTCTCCCCCCCCCTAAG 7300
Qy 6961 AAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGTGAGGACTCCATA 7020
Db 7301 AAAACCCCGACGCCTCCCCCAAGGAGACGCCGGACAGTGGGTCTGAGTGAGAGCTCCATA 7360
Qy 7021 GGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCCCCAAGCGGCGAT 7080
Db 7361 GCAGATGCCCTACAACAGCTGGCCATCAAGTCCTTTGGCCAGCCCCCCCCAAGCGGCGAT 7420
Qy 7081 TCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACGCCTCCTGATGAG 7140
Db 7421 TCAGGCCTTTCCACGGGGGCGGACGCAGCCGATTCCGGCAGTCGGACGCCCCCCGATGAG 7480
Qy 7141 TTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCTCGAGGGGGAGCTTGGA 7200
Db 7481 TTGGCCCTTTCGGAGACAGGTTCCATCTCCTCCATGCCCCCTCTCGAGGGGGAGCCTGGA 7540
Qy 7201 GATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCCAGGGGGGGGTGGCA 7260

```
Db       7541 GATCCAGACTTGGAGCCTGAGCAGGTAGAGCTTCAACCTCCCCCCCAGGGGGGGGTGGTA 7600

Qy       7261 GCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGACGACTCCGTCGTG 7320

Db       7601 ACCCCCGGCTCAGGCTCGGGGTCTTGGTCTACTTGCTCCGAGGAGGACGACTCCGTCGTG 7660

Qy       7321 TGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGTAGTCCCGAAGAG 7380

Db       7661 TGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGTAGCCCCGAAGAG 7720

Qy       7381 GAGAAGTTACCGATTAACCCCTTGAGCAACTCCCTGTTGCGATATCACAACAAGGTGTAC 7440

Db       7721 GAAAAGTTGCCAATTGGCCCCTTGAGCAACTCCCTGTTGCGATATCACAACAAGGTGTAC 7780

Qy       7441 TGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTTGATAGGATGCAA 7500

Db       7781 TGTACCACATCAAAGAGCGCCTCATTAAGGGCTAAAAAGGTAACTTTTGATAGGATGCAA 7840

Qy       7501 GTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCGGCCTCCAAGGTC 7560

Db       7841 GCGCTCGACGCTCATTATGACTCAGTCTTGAAGGACATTAAGCTAGCGGCCTCCAAGGTC 7900

Qy       7561 ACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCCCATTCTGCAAGA 7620

Db       7901 ACCGCAAGGCTTCTCACTTTAGAGGAGGCCTGCCAGTTAACTCCACCCCACTCTGCAAGA 7960

Qy       7621 TCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGGGCCGTTAACCAC 7680

Db       7961 TCCAAGTATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGAGCCGTTAACCAC 8020

Qy       7681 ATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATTCCCACAACCATT 7740

Db       8021 ATCAAGTCCGTGTGGAAGGACCTCCTGGAAGACACACAAACACCAATTCCTACAACCATC 8080

Qy       7741 ATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAGAAAGCAGCTCGC 7800

Db       8081 ATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGTAAGAAAGCAGCTCGC 8140

Qy       7801 CTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCCCTTTATGACATT 7860

Db       8141 CTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAAATGGCCCTTTATGATATC 8200

Qy       7861 ACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAGTATTCCCCCGCT 7920

Db       8201 ACACAAAAGCTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAGTACTCCCCCGCT 8260

Qy       7921 CAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCTATGGGTTTTTCG 7980

Db       8261 CAGCGGGTGGAGTTTCTCTTGAAGGCATGGGCGGAAAAGAAAGACCCTATGGGTTTTTCG 8320

Qy       7981 TATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGGACTGAGGAGTCC 8040

Db       8321 TATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGGACTGAGGAGTCC 8380

Qy       8041 ATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACACTCGCTAACTGAG 8100

Db       8381 ATATATCGGGCTTGTTCCTTGCCCGAGGAGGCCCACACTGCCATACACTCACTGACTGAG 8440

Qy       8101 AGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGCGGGTACAGGCGT 8160

Db       8441 AGACTTTACGTGGGAGGGCCCATGTTCAACAGCAAGGGCCAGACCTGCGGGTACAGGCGT 8500

Qy       8161 TGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACATGCTACGTGAAA 8220

Db       8501 TGCCGCGCCAGCGGGGTGCTTACCACTAGCATGGGGAACACCATCACATGCTATGTGAAA 8560

Qy       8221 GCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTGGTATGCGGCGAT 8280

Db       8561 GCCTTAGCGGCCTGTAAGGCTGCAGGGATAATTGCGCCCACAATGCTGGTATGCGGCGAT 8620

Qy       8281 GACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGGAACCTGAGAGCC 8340

Db       8621 GACTTGGTTGTCATCTCAGAGAGCCAGGGGACCGAGGAGGACGAGCGGAACCTGAGAGCC 8680

Qy       8341 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCCAGACCGGAGTAT 8400

Db       8681 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCCAGACCGGAATAT 8740

Qy       8401 GATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTGGGCCCACAAGGC 8460

Db       8741 GACCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGTTGGGCCCACAAGGC 8800

Qy       8461 CGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGGGCTGCCTGGGAA 8520

Db       8801 CGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGGGCTGCCTGGGAA 8860

Qy       8521 ACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAGTACGCCCCGACC 8580

Db       8861 ACAGTTAGACACTCCCCTGTCAATTCATGGCTAGGAAACATCATCCAGTACGCCCCAACC 8920

Qy       8581 ATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATGGCTCAAGACACG 8640

Db       8921 ATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATGGCCCAAGATACT 8980

Qy       8641 CTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTGAGTCCCTTGGAC 8700

Db       8981 CTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTGAGTCCCTTGGAC 9040

Qy       8701 CTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTGCACACATACACT 8760

Db       9041 CTCCCAGCCATAATTGAAAGGTTACACGGGCTTGACGCTTTCTCTCTGCACACATACACT 9100

Qy       8761 CCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGA 8820

Db       9101 CCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGA 9160

Qy       8821 GCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGTGGGGGGAGAGCG 8880

Db       9161 GCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGTGGGGGGAGAGCG 9220

Qy       8881 GCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 8940

Db       9221 GCCGTTTGCGGCCGATATCTCTTCAACTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 9280

Qy       8941 TTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGCGCCGGCGGGGGC 9000

Db       9281 TTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACTGTCGGCGCCGGCGGGGGC 9340

Qy       9001 GACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTTGGCCTACTCCTA 9060

Db       9341 GACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTACTCCTTGGCCTACTCCTA 9400

Qy       9061 CTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCGG 9099

Db       9401 CTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCGG 9439
```

```
RESULT 9
AB047641
LOCUS       AB047641               9659 bp    RNA     linear   VRL 04-AUG-2001
DEFINITION  Hepatitis C virus gene for polyprotein, complete cds, clone:JCH-2.
ACCESSION   AB047641
VERSION     AB047641.1  GI:13122265
KEYWORDS    .
SOURCE      Hepatitis C virus
  ORGANISM  Hepatitis C virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepacivirus.
REFERENCE   1
  AUTHORS   Kato,T., Furusaka,A., Miyamoto,M., Date,T., Yasui,K., Hiramoto,J.,
            Nagayama,K., Tanaka,T. and Wakita,T.
  TITLE     Sequence analysis of hepatitis C virus isolated from a fulminant
            hepatitis patient
  JOURNAL   J. Med. Virol. 64 (3), 334-339 (2001)
  MEDLINE   21316767
```

```
   PUBMED   11424123
REFERENCE   2  (bases 1 to 9659)
  AUTHORS   Kato,T., Wakita,T. and Furusaka,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-AUG-2000) Takanobu Kato, The Tokyo Metropolitan
            Institute for Neuroscience, Department of Microbiology; Musashidai
            2-6, Fuchu, Tokyo 183-8526, Japan (E-mail:takato@tmin.ac.jp,
            Tel:81-423-25-3881(ex.4605), Fax:81-423-21-8678)
FEATURES             Location/Qualifiers
     source          1. .9659
                     /organism="Hepatitis C virus"
                     /mol_type="genomic RNA"
                     /db_xref="taxon:11103"
                     /clone="JCH-2"
                     /note="isolated from chronic hepatitis patient/genotype
                     2a"
     CDS             341. .9442
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="BAB32874.1"
                     /db_xref="GI:13122266"
                     /translation="MSTTPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRL
                     GVRATRKTSERSQPRGRRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRG
                     SRPSWGPNDPRHRSRNVGKVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLED
                     GVNFATGNLPGCSFSIFLLALLSCITTPVSGAQVRNTSSSYMVANDCANDSITWQLDA
                     AVLHVPGCVPCEKVGDTSRCWIPVSPNVAVRQPGALTRGLRTHIDMVVMSATLCSALY
                     VGDLCGGVMLAAQMFIVSPQRHWFVQECNCSIYPGHVSGHRMAWDMMMNWSPTATMIL
                     AYVMRVPEVIIDIISGAHWGVMFGLAYFSMQGAWAKVVVILLLAAGVDAGTYAVGGST
                     AHTTKGIASFFSPGARQNIQLINTNGSWHINRTALNCNESLNTGFLAALFYANRFNSS
                     GCPERLSACRNIEDFRIGWGTLQYENNVTNSEDMRPYCWYYPPRQCGIVPARSVCGPV
                     YCFTPSPVVVGTTDRLGVPTYTWGENETDVFLLNSTRPPSGSWFGCTWMNSTGFTKTC
                     GAPPCRIRADFNASADLLCPTDCFRKHPDAAYFRCGSGPWLTPRCLVDYPYRLWHYPC
                     TVNYTIFKIRMYVGGVEHRLAAACNFTRGDRCDLEDRDRSQLSPLLHSTTEWAILPCT
                     YSDLPALSTGLLHLHQNIVDVQYMYGLSPALTKYIVRWEWVVLLFLLLADARVCACVW
                     MLILLGQAEAALEKLVVLHAASAASCNGFLYFVIFFVAAWYIKGRAVPLATYSLTGLW
                     SFCLLLLALPQQAYAYDTSAHGQIGATLLILITLFTLTPAYKTLVSRFLWWLCYLLTL
                     GEAMVQEWAPPLQVRGGRDGIIWAAAIFCPGVVFDITKWLLALLGPAYLLRDALTRVP
                     YFVRAHALLRICTMVRHLAGGRYVQMALLALGRWTGTYIYDHLSPMSDWAANGLRDLA
                     VAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPADGYTSKGWSLL
                     APITAYAQQTRGLLGAIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSIAGVLWTVYHG
                     AGNKTLAGSRGPVTQMYSSAEGDLVEWPSPPGTKSLEPCTCGAVDLYLVTRNADVIPA
                     RRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGIFRAAVCSRGVAKSIDFIPVET
                     LDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNPSV
                     AATLGFGAYLSKAHGINPNIRTGVRTVTTGAAITYSTYGKFLADGGCSGGAYDIIICD
                     ECHSVDATTILGIGTVLDQAETAGVRSTVLATATPPGSVTTPHPNIEEVALGQGEGEIP
                     FYGRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNAVAYYRGLDVSIIPTQGDV
                     VVVATDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQTVPQDAVSRSQRRGR
                     TGRGRLGIYRYVSTGERASGMFDSVVLSECYDSGAAWYELTPAETTVRLRAYFNTPGL
                     PVCQDHLEFWEAVFTGLTHIDAHFLSQTKQAGDNFAYLTAYQATVCARAKAPPSSWDV
                     MWKCLTRLKPTLVGPTPLLYRLGPVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAG
                     GVLAAVAAYCLATGCVSIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQ
                     RIAEMLKSKIQGLLQQASKQAQDIQPAVQASWPKVEQFWAKHMWNFTSGIQYLAGLST
                     LPGNPAVASMMAFSAALTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAA
                     VGSIGLGKVLVDILAGYGAGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGV
                     ICAAILRRHVGPGEGAVQWMNRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTIT
                     SLLRRLHNWITEGCPIPCSGSWLRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQ
                     KGYKGVWAGTGIMTTRCPCGANISGNVRLGSMRITGPKTCMNTWQGTFPINCYTGGQC
                     VPKPAPNFKTAIWRVAASEYAEVTQHGSYSYITGLTTDNLKVPCQLPSPEFFSWVDGV
                     QIHRFAPTPKPFFRDEVSFCVGLNSFVVGSQLPCDPEPDTDVLMSMLTDPSHITAETA
                     ARRLARGSPPSEASSSASQLSAPSLRATCTTHGKTHDVDMVDANLFMGGDVTRIESES
                     KVVVLDSLDPMVEEEGDLEPSIPSEYMLPKKKFPPALPPWARPDYNPPLVESRKKPDY
                     QPPTVAGCALPPPKKTPTPPPRRRRTVGLSESTIGDALQQLAIKAFGQPPPSGDSGLS
                     TGADAADSGSRTPPDESALSETGSISSMPPLEGEPGDPDLEPEQVERQPPPQGGEVAP
                     GSDSGSWSTCSGEDDSVVCCSMSYSWTGALITPCSPEEEKLPINPLSNSLLRYHNKVY
                     CTTSKSASLRAKKVTFDRMQVLDAYYDSVLKDIKLAASKVTAKLLTLEEACQLTPPHS
                     ARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSQTPIPTTIMAKNEVFCVDPTKGGK
                     KAARLIVYPDLGVRVCEKMALYDVTQKLPQAVMGASYGFQYSPAQRVEFLLKAWAEKK
                     DPMGFSYDTRCFDTTVTERDIRTEESIYRACSLPEEAHTAIHSLTERLYVGGPMFNSK
                     GQSCGYRRCRASGVLTTSMGNTITCYAKALAACKAAGIVAPTMLVCGDDLVVISESQG
                     TEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQGRRRYYLTR
                     DPTTPIARAAWETVRHSPANSWLGNIIQYAPTIWARMVLMTHFFSILMAQDTLDQNLN
                     FEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYSPHELTRVASALRKLGAPPLRAWKS
                     RARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGGDI
                     YHSVSRARPRLLLLSLLLLSVGVGLFLLPAR"
ORIGIN

  Query Match           88.2%;  Score 8025.4;  DB 14;  Length 9659;
  Best Local Similarity  92.6%;  Pred. No. 0;
  Matches 8428;  Conservative    0;  Mismatches  671;  Indels    0;  Gaps    0;

Qy          1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 60
              ||||||||| |||||||||||||||||||||||||||||||||| ||||||||||||||
Db        341 ATGAGCACAACTCCTAAACCTCAAAGAAAAACCAAAAGAAACACTAACCGTCGCCCACAA 400

Qy         61 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 460

Qy        121 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 520

Qy        181 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 240
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        521 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAGTCCTGGGGAAAACCAGGA 580

Qy        241 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 300
              ||||| |||||||| || |||||||||||||||||||||||||||||||||||||||||
Db        581 TACCCTTGGCCCCTGTATGGGAATGAGGGGCTCGGCTGGGCAGGATGGCTCCTGTCCCCC 640

Qy        301 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 360
              |||||||| |||||||||||||||||||||||||| ||||||||||||||||||||||
Db        641 CGAGGTTCTCGTCCCTCTTGGGGCCCCAATGACCCCCGGCACAGGTCGCGCAACGTGGGT 700

Qy        361 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 420
              || |||||||| ||||||||||||||||||||||||||||||||||||||| |||||
Db        701 AAAGTCATCGACACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCCGTCGTA 760

Qy        421 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 480
              |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db        761 GGCGCCCCGCTTGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 820

Qy        481 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 540
              |||||||||||||||||||||||| || |||||||| ||||||||||||||||||||||
Db        821 GGGGTTAATTTTGCAACAGGGAACTTGCCTGGTTGCTCTTTTTCTATCTTCTTGCTGGCT 880

Qy        541 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 600
              |||||||||||||||||||| |||||||| |  ||| |||||| ||||| |||| || ||
Db        881 CTGCTGTCCTGCATCACCACACCGGTCTCTGGCGCCCAAGTGAGGAACACCAGTTCCAGC 940

Qy        601 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 660
              || |||||| | || |||||  | |||||||||| |||||||||| | |||||  |||
Db        941 TATATGGTGGCCAATGACTGTGCTAATGACAGCATCACCTGGCAGCTCGACGCTGCGGTC 1000

Qy        661 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 720
              |||||||||||||||||||||||||||||||| |||||| |  |||||| ||||||||||
Db       1001 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAGGTGGGGGACACATCTCGGTGCTGGATA 1060

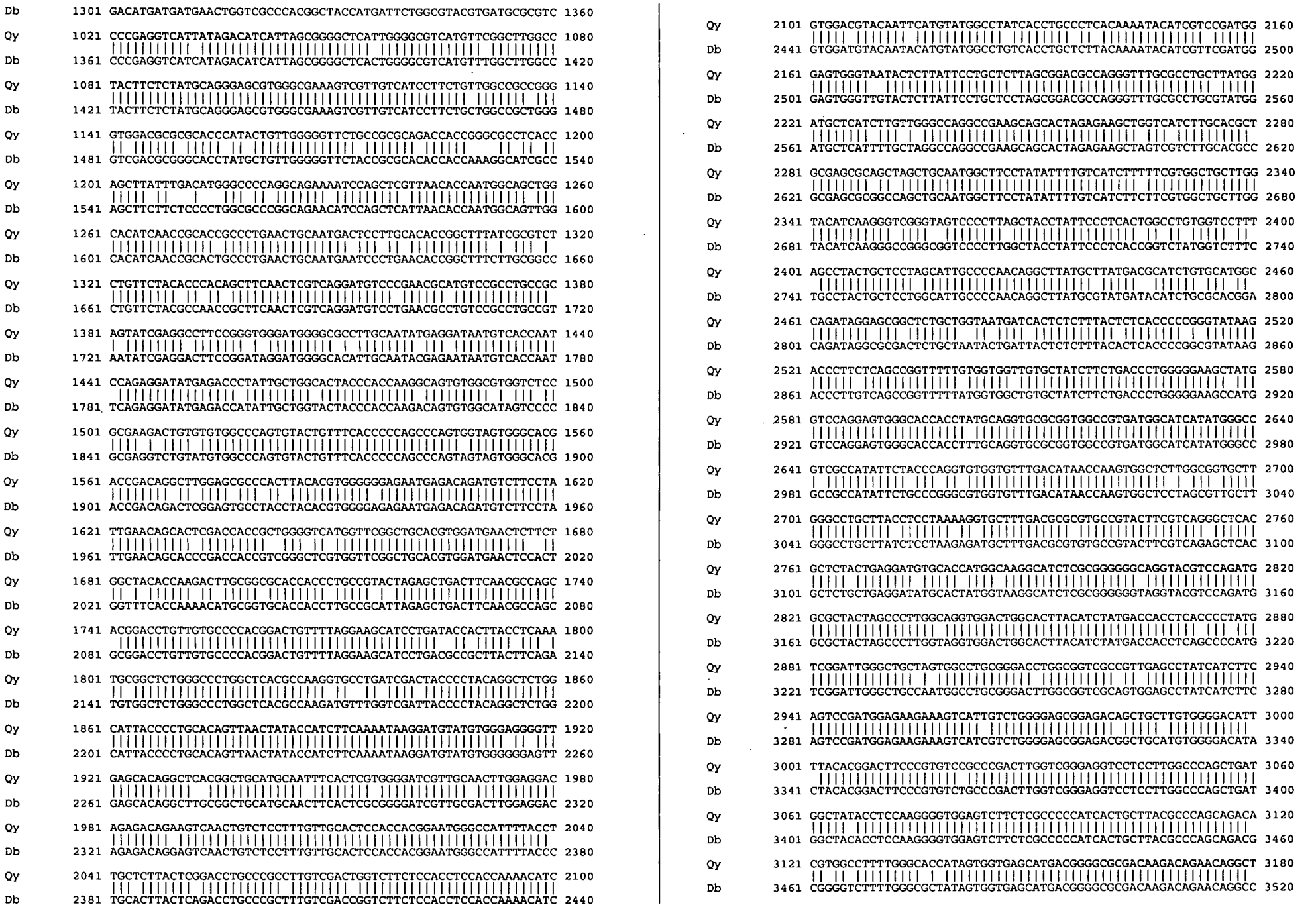
Qy        721 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 780
              ||||| ||||| || |||||||||| || |||||| |||||||||  ||||||||||||
Db       1061 CCGGTTTCACCAAACGTGGCCGTGCGGCAGCCCGGTGCCCTCACGCGAGGCTTGCGGACG 1120

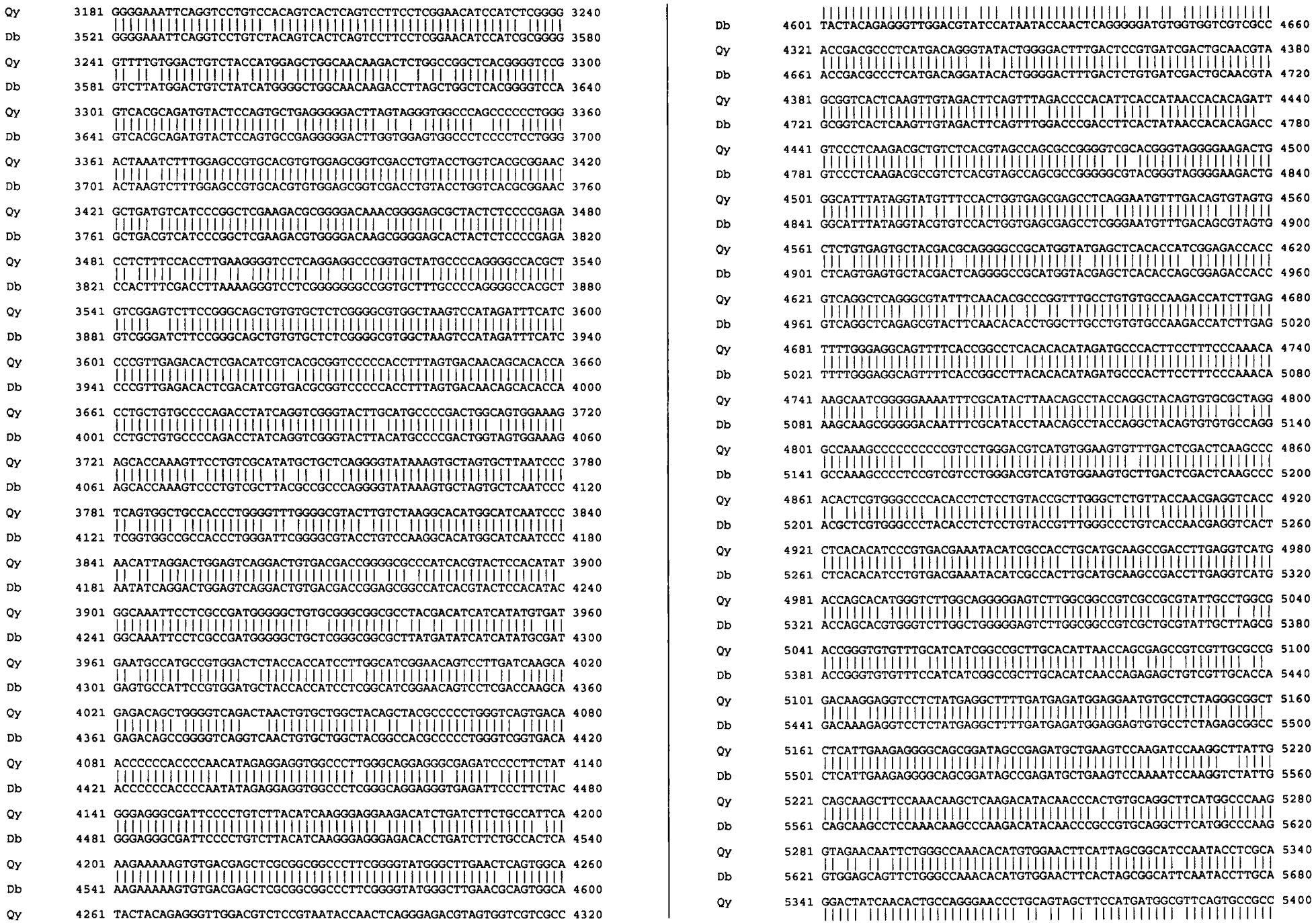
Qy        781 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 840
              ||||| |||||||| || |||||||||||||||||||||| || ||||||||||||||||
Db       1121 CACATAGACATGGTGGTAATGTCCGCCACGCTCTGCTCCGCTCTCTACGTGGGGGACCTT 1180

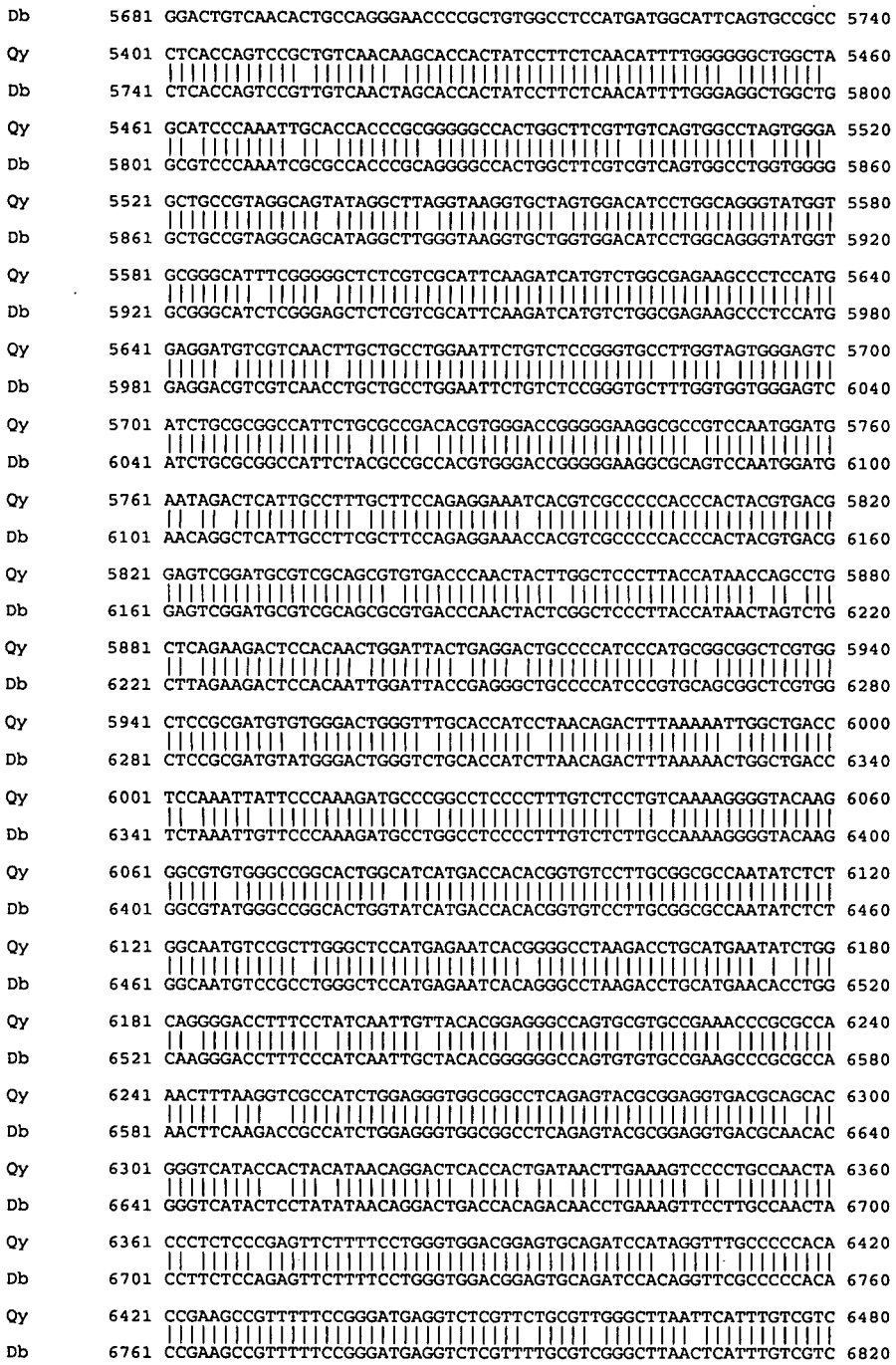
Qy        841 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 900
              || |||||||||||||||||||| |||||||||||||||| |||| ||||||||||
Db       1181 TGTGGTGGGGTGATGCTCGCAGCCCAGATGTTCATTGTCTCGCCACAGCGCCACTGGTTT 1240

Qy        901 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 960
              || ||||| |||||||||||||||||||| |||  | ||| ||||||||||||||||||
Db       1241 GTGCAAGAATGCAATTGCTCCATCTACCCCGGTCACGTCAGTGGACACCGCATGGCATGG 1300

Qy        961 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1020
              |||||||||||||||||||||||||||||||||||||||  ||||||||| |||||| |||
```


Db 1301 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATTCTGGCGTACGTGATGCGCGTC 1360
Qy 1021 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1080
Db 1361 CCCGAGGTCATCATAGACATCATTAGCGGGGCTCACTGGGGCGTCATGTTTGGCTTGGCC 1420
Qy 1081 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1140
Db 1421 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGCTGGCCGCTGGG 1480
Qy 1141 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1200
Db 1481 GTCGACGCGGGCACCTATGCTGTTGGGGGTTCTACCGCGCACACCACCAAAGGCATCGCC 1540
Qy 1201 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1260
Db 1541 AGCTTCTTCTCCCCTGGCGCCCGGCAGAACATCCAGCTCATTAACACCAATGGCAGTTGG 1600
Qy 1261 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1320
Db 1601 CACATCAACCGCACTGCCCTGAACTGCAATGAATCCCTGAACACCGGCTTTCTTGCGGCC 1660
Qy 1321 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1380
Db 1661 CTGTTCTACGCCAACCGCTTCAACTCGTCAGGATGTCCTGAACGCCTGTCCGCCTGCCGT 1720
Qy 1381 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1440
Db 1721 AATATCGAGGACTTCCGGATAGGATGGGGCACATTGCAATACGAGAATAATGTCACCAAT 1780
Qy 1441 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1500
Db 1781 TCAGAGGATATGAGACCATATTGCTGGTACTACCCACCAAGACAGTGTGGCATAGTCCCC 1840
Qy 1501 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1560
Db 1841 GCGAGGTCTGTATGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTAGTAGTGGGCACG 1900
Qy 1561 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1620
Db 1901 ACCGACAGACTCGGAGTGCCTACCTACACGTGGGGAGAGAATGAGACAGATGTCTTCCTA 1960
Qy 1621 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 1680
Db 1961 TTGAACAGCACCCGACCACCGTCGGGCTCGTGGTTCGGCTGCACGTGGATGAACTCCACT 2020
Qy 1681 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 1740
Db 2021 GGTTTCACCAAAACATGCGGTGCACCACCTTGCCGCATTAGAGCTGACTTCAACGCCAGC 2080
Qy 1741 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 1800
Db 2081 GCGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGACGCCGCTTACTTCAGA 2140
Qy 1801 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 1860
Db 2141 TGTGGCTCTGGGCCCTGGCTCACGCCAAGATGTTTGGTCGATTACCCCTACAGGCTCTGG 2200
Qy 1861 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 1920
Db 2201 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGGGGAGTT 2260
Qy 1921 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 1980
Db 2261 GAGCACAGGCTTGCGGCTGCATGCAACTTCACTCGCGGGGATCGTTGCGACTTGGAGGAC 2320
Qy 1981 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2040
Db 2321 AGAGACAGGAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCC 2380
Qy 2041 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2100
Db 2381 TGCACTTACTCAGACCTGCCCGCTTTGTCGACCGGTCTTCTCCACCTCCACCAAAACATC 2440
Qy 2101 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2160
Db 2441 GTGGATGTACAATACATGTATGGCCTGTCACCTGCTCTTACAAAATACATCGTTCGATGG 2500
Qy 2161 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2220
Db 2501 GAGTGGGTTGTACTCTTATTCCTGCTCCTAGCGGACGCCAGGGTTTGCGCCTGCGTATGG 2560
Qy 2221 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2280
Db 2561 ATGCTCATTTTGCTAGGCCAGGCCGAAGCAGCACTAGAGAAGCTAGTCGTCTTGCACGCC 2620
Qy 2281 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATATTTTGTCATCTTTTTCGTGGCTGCTTGG 2340
Db 2621 GCGAGCGCGGCCAGCTGCAATGGCTTCCTATATTTTGTCATCTTCTTCGTGGCTGCTTGG 2680
Qy 2341 TACATCAAGGGTCGGGTAGTCCCCTTAGCTACCTATTCCCTCACTGGCCTGTGGTCCTTT 2400
Db 2681 TACATCAAGGGCCGGGCGGTCCCCTTGGCTACCTATTCCCTCACCGGTCTATGGTCTTTC 2740
Qy 2401 AGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTGCATGGC 2460
Db 2741 TGCCTACTGCTCCTGGCATTGCCCCAACAGGCTTATGCGTATGATACATCTGCGCACGGA 2800
Qy 2461 CAGATAGGAGCGGCTCTGCTGGTAATGATCACTCTCTTTACTCTCACCCCCGGGTATAAG 2520
Db 2801 CAGATAGGCGCGACTCTGCTAATACTGATTACTCTCTTTACACTCACCCCGGCGTATAAG 2860
Qy 2521 ACCCTTCTCAGCCGGTTTTTGTGGTGGTTGTGCTATCTTCTGACCCTGGGGGAAGCTATG 2580
Db 2861 ACCCTTGTCAGCCGGTTTTTATGGTGGCTGTGCTATCTTCTGACCCTGGGGGAAGCCATG 2920
Qy 2581 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGCATCATATGGGCC 2640
Db 2921 GTCCAGGAGTGGGCACCACCTTTGCAGGTGCGCGGTGGCCGTGATGGCATCATATGGGCC 2980
Qy 2641 GTCGCCATATTCTACCCAGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 2700
Db 2981 GCCGCCATATTCTGCCCGGGCGTGGTGTTTGACATAACCAAGTGGCTCCTAGCGTTGCTT 3040
Qy 2701 GGGCCTGCTTACCTCCTAAAAGGTGCTTTGACGCGCGTGCCGTACTTCGTCAGGGCTCAC 2760
Db 3041 GGGCCTGCTTATCTCCTAAGAGATGCTTTGACGCGTGTGCCGTACTTCGTCAGAGCTCAC 3100
Qy 2761 GCTCTACTGAGGATGTGCACCATGGCAAGGCATCTCGCGGGGGGCAGGTACGTCCAGATG 2820
Db 3101 GCTCTGCTGAGGATATGCACTATGGTAAGGCATCTCGCGGGGGGTAGGTACGTCCAGATG 3160
Qy 2821 GCGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 2880
Db 3161 GCGCTACTAGCCCTTGGTAGGTGGACTGGCACTTACATCTATGACCACCTCAGCCCCATG 3220
Qy 2881 TCGGATTGGGCTGCTAGTGGCCTGCGGGACCTGGCGGTCGCCGTTGAGCCTATCATCTTC 2940
Db 3221 TCGGATTGGGCTGCCAATGGCCTGCGGGACTTGGCGGTCGCAGTGGAGCCTATCATCTTC 3280
Qy 2941 AGTCCGATGGAGAAGAAAGTCATTGTCTGGGGAGCGGAGACAGCTGCTTGTGGGGACATT 3000
Db 3281 AGTCCGATGGAGAAGAAAGTCATCGTCTGGGGAGCGGAGACGGCTGCATGTGGGGACATA 3340
Qy 3001 TTACACGGACTTCCCGTGTCCGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3060
Db 3341 CTACACGGACTTCCCGTGTCTGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3400
Qy 3061 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3120
Db 3401 GGCTACACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACG 3460
Qy 3121 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCT 3180
Db 3461 CGGGGTCTTTTGGGCGCTATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCC 3520




Db 5681 GGACTGTCAACACTGCCAGGGAACCCCGCTGTGGCCTCCATGATGGCATTCAGTGCCGCC 5740
Qy 5401 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5460
Db 5741 CTCACCAGTCCGTTGTCAACTAGCACCACTATCCTTCTCAACATTTTGGGAGGCTGGCTG 5800
Qy 5461 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5520
Db 5801 GCGTCCCAAATCGCGCCACCCGCAGGGGCCACTGGCTTCGTCGTCAGTGGCCTGGTGGGG 5860
Qy 5521 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5580
Db 5861 GCTGCCGTAGGCAGCATAGGCTTGGGTAAGGTGCTGGTGGACATCCTGGCAGGGTATGGT 5920
Qy 5581 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5640
Db 5921 GCGGGCATCTCGGGAGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5980
Qy 5641 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 5700
Db 5981 GAGGACGTCGTCAACCTGCTGCCTGGAATTCTGTCTCCGGGTGCTTTGGTGGTGGGAGTC 6040
Qy 5701 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 5760
Db 6041 ATCTGCGCGGCCATTCTACGCCGCCACGTGGGACCGGGGGAAGGCGCAGTCCAATGGATG 6100
Qy 5761 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 5820
Db 6101 AACAGGCTCATTGCCTTCGCTTCCAGAGGAAACCACGTCGCCCCCACCCACTACGTGACG 6160
Qy 5821 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 5880
Db 6161 GAGTCGGATGCGTCGCAGCGCGTGACCCAACTACTCGGCTCCCTTACCATAACTAGTCTG 6220
Qy 5881 CTCAGAAGACTCCACAACTGGATTACTGAGGACTGCCCCATCCCATGCGGCGGCTCGTGG 5940
Db 6221 CTTAGAAGACTCCACAATTGGATTACCGAGGGCTGCCCCATCCCGTGCAGCGGCTCGTGG 6280
Qy 5941 CTCCGCGATGTGTGGGACTGGGTTTGCACCATCCTAACAGACTTTAAAAATTGGCTGACC 6000
Db 6281 CTCCGCGATGTATGGGACTGGGTCTGCACCATCTTAACAGACTTTAAAAACTGGCTGACC 6340
Qy 6001 TCCAAATTATTCCCAAAGATGCCCGGCCTCCCCTTTGTCTCCTGTCAAAAGGGGTACAAG 6060
Db 6341 TCTAAATTGTTCCCAAAGATGCCTGGCCTCCCCTTTGTCTCTTGCCAAAAGGGGTACAAG 6400
Qy 6061 GGCGTGTGGGCCGGCACTGGCATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6120
Db 6401 GGCGTATGGGCCGGCACTGGTATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6460
Qy 6121 GGCAATGTCCGCTTGGGCTCCATGAGAATCACGGGGCCTAAGACCTGCATGAATATCTGG 6180
Db 6461 GGCAATGTCCGCCTGGGCTCCATGAGAATCACAGGGCCTAAGACCTGCATGAACACCTGG 6520
Qy 6181 CAGGGGACCTTTCCTATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCGCCA 6240
Db 6521 CAAGGGACCTTTCCCATCAATTGCTACACGGGGGGCCAGTGTGTGCCGAAGCCCGCGCCA 6580
Qy 6241 AACTTTAAGGTCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6300
Db 6581 AACTTCAAGACCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAACAC 6640
Qy 6301 GGGTCATACCACTACATAACAGGACTCACCACTGATAACTTGAAAGTCCCCTGCCAACTA 6360
Db 6641 GGGTCATACTCCTATATAACAGGACTGACCACAGACAACCTGAAAGTTCCTTGCCAACTA 6700
Qy 6361 CCCTCTCCCGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCACA 6420
Db 6701 CCTTCTCCAGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCACAGGTTCGCCCCCACA 6760
Qy 6421 CCGAAGCCGTTTTTTCCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6480
Db 6761 CCGAAGCCGTTTTTTCCGGGATGAGGTCTCGTTTTGCGTCGGGCTTAACTCATTTGTCGTC 6820
Qy 6481 GGGTCCCAGCTTCCTTGCGACCCTGAACCCGACACAGACGTATTGATGTCCATGCTAACA 6540
Db 6821 GGGTCTCAGCTTCCTTGTGACCCTGAACCCGACACAGACGTATTAATGTCCATGCTAACA 6880
Qy 6541 GATCCATCTCATATCACGGCGGAGACTGCAGCGCGGCGTTTAGCGCGGGGGTCACCCCCA 6600
Db 6881 GATCCATCTCATATCACGGCGGAGACCGCAGCGCGGCGCTTGGCGCGGGGGTCACCCCCA 6940
Qy 6601 TCCGAGGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 6660
Db 6941 TCTGAGGCAAGCTCCTCAGCGAGCCAGTTATCGGCGCCATCGCTGCGAGCCACCTGCACC 7000
Qy 6661 ACCCACGGCAAAGCCTATGATGTGGACATGGTGGATGCTAACCTGTTCATGGGGGGCGAT 6720
Db 7001 ACCCACGGCAAGACCCATGATGTGGACATGGTGGATGCCAACCTATTCATGGGGGGCGAT 7060
Qy 6721 GTGACTCGGATAGAGTCTGGGTCCAAAGTGGTCGTTCTGGACTCTCTCGACCCAATGGTC 6780
Db 7061 GTCACTCGGATAGAGTCCGAGTCCAAAGTGGTCGTTCTGGACTCTCTCGATCCAATGGTC 7120
Qy 6781 GAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCAGAATACATGCTCCCCAAGAAGAGG 6840
Db 7121 GAAGAAGAGGGCGACCTCGAGCCTTCGATACCATCAGAGTACATGCTCCCCAAAAAGAAG 7180
Qy 6841 TTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCGCTTGTGGAATCG 6900
Db 7181 TTTCCACCAGCCTTACCGCCTTGGGCGCGGCCAGACTACAACCCACCGCTTGTGGAATCA 7240
Qy 6901 TGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTCCCTCCTCCTAGG 6960
Db 7241 CGGAAGAAGCCAGACTATCAACCGCCCACTGTTGCGGGCTGTGCTCTCCCCCCCCCTAAG 7300
Qy 6961 AAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGTGAGGACTCCATA 7020
Db 7301 AAGACCCCGACGCCCCCCCCAAGGAGGCGCCGGACAGTGGGTCTGAGTGAGAGCACCATA 7360
Qy 7021 GGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCCCCCAAGCGGCGAT 7080
Db 7361 GGGGATGCCCTACAACAGCTGGCTATTAAGGCTTTTGGCCAGCCCCCCCCCAAGCGGAGAT 7420
Qy 7081 TCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACGCCTCCTGATGAG 7140
Db 7421 TCAGGCCTGTCCACGGGGGCGGACGCCGCCGACTCCGGCAGTCGGACGCCCCCTGATGAG 7480
Qy 7141 TTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCCTCGAGGGGGAGCTTGGA 7200
Db 7481 TCGGCTCTTTCGGAGACAGGTTCCATCTCCTCCATGCCCCCCCCTCGAGGGGGAGCCTGGA 7540
Qy 7201 GATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCCAGGGGGGGGTGGCA 7260
Db 7541 GATCCAGACCTGGAGCCTGAGCAGGTAGAGCGTCAACCTCCCCCCCAGGGGGGGGAAGTA 7600
Qy 7261 GCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGACGACTCCGTCGTG 7320
Db 7601 GCTCCCGGCTCGGACTCGGGGTCCTGGTCCACTTGCTCCGGCGAGGACGACTCCGTCGTG 7660
Qy 7321 TGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGTAGTCCCGAAGAG 7380
Db 7661 TGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGTAGCCCCGAAGAG 7720
Qy 7381 GAGAAGTTACCGATTAACCCCTTGAGCAACTCCCTGTTGCGATATCACAACAAGGTGTAC 7440
Db 7721 GAAAAGTTACCGATTAACCCTTTGAGCAACTCGCTGTTGCGATACCACAACAAGGTGTAC 7780
Qy 7441 TGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTTGATAGGATGCAA 7500
Db 7781 TGTACCACATCAAAGAGCGCCTCACTGAGGGCTAAAAAGGTAACTTTTGATAGGATGCAA 7840
Qy 7501 GTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCGGCCTCCAAGGTC 7560
Db 7841 GTGCTCGACGCCTATTACGACTCAGTCTTAAAGGACATCAAGCTAGCGGCCTCCAAGGTC 7900

```
Qy       7561 ACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCCCATTCTGCAAGA 7620
             || |||| |||||||||| |||||||||| |||||||||||||||||||| |||||||||
Db       7901 ACTGCAAAGCTCCTCACCTTGGAGGAGGCATGCCAGTTAACCCCACCCCACTCTGCAAGA 7960

Qy       7621 TCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGGGCCGTTAACCAC 7680
             || || || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7961 TCCAAGTACGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGGGCCGTTAACCAC 8020

Qy       7681 ATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATTCCCACAACCATT 7740
             |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       8021 ATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCACAAACACCAATTCCCACAACCATC 8080

Qy       7741 ATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAGAAAGCAGCTCGC 7800
             |||||||||||||||||||||||||||||||||| ||||| ||||||||||||||||||
Db       8081 ATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAAGGGGGTAAGAAAGCAGCTCGC 8140

Qy       7801 CTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCCCTTTATGACATT 7860
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||   |
Db       8141 CTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCCCTTTATGATGTC 8200

Qy       7861 ACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAGTATTCCCCCGCT 7920
             |||||||||| |||||||||| |||||||||||||||||||||||||||||||| ||||||||
Db       8201 ACACAAAAACTCCCTCAGGCGGTAATGGGGGCTTCTTATGGATTCCAGTACTCCCCCGCC 8260

Qy       7921 CAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCTATGGGTTTTTCG 7980
             ||||||| |||||||||||||| |||||||||||||||| ||||||||||||||||||||
Db       8261 CAGCGGGTGGAGTTTCTCTTGAAGGCATGGGCGGAAAAGAAAGACCCTATGGGTTTTTCG 8320

Qy       7981 TATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGGACTGAGGAGTCC 8040
             |||||||||||||||||||| |||||||||| ||||||||||||||||||||||||||||
Db       8321 TATGATACCCGATGCTTTGACACAACCGTCACCGAGAGAGACATCAGGACTGAGGAGTCC 8380

Qy       8041 ATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACACTCGCTAACTGAG 8100
             ||||| |||||||| ||| ||||||||||||||||||||||||||||||||||||||||||
Db       8381 ATATACCGGGCCTGTTCCCTGCCCGAGGAGGCCCACACTGCCATACACTCGCTAACTGAG 8440

Qy       8101 AGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGCGGGTACAGGCGT 8160
             |||||||||||||||||||||||| |||||||||||||  ||||||||||||||||||||
Db       8441 AGACTTTACGTGGGAGGGCCTATGTTTAACAGCAAGGGCCAGTCCTGCGGGTACAGGCGT 8500

Qy       8161 TGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACATGCTACGTGAAA 8220
             |||||||||| |||||||||||||||||||||||||||||||| |||||||  ||| |||
Db       8501 TGCCGCGCCAGTGGGGTGCTCACCACTAGCATGGGGAACACCATCACGTGCTACGCAAAA 8560

Qy       8221 GCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTGGTATGCGGCGAT 8280
             ||| ||||||| || || |||||||||||| | ||||||||||||||||||||||||||||
Db       8561 GCCCTAGCGGCCTGCAAGGCTGCAGGGATAGTTGCGCCCACAATGCTGGTATGCGGCGAC 8620

Qy       8281 GACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGGAACCTGAGAGCC 8340
             |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db       8621 GACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGATGAGCGGAACCTGAGAGCC 8680

Qy       8341 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCCAGACCGGAGTAT 8400
             |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||| |||
Db       8681 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGCGACCCCCCCAGACCGGAATAT 8740

Qy       8401 GATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTGGGCCCACAAGGC 8460
             ||  ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db       8741 GACTTGGAGCTGATAACATCTTGTTCCTCAAATGTGTCTGTGGCGTTGGGCCCACAAGGC 8800

Qy       8461 CGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGGGCTGCCTGGGAA 8520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8801 CGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGGGCTGCCTGGGAA 8860

Qy       8521 ACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAGTACGCCCCGACC 8580
             |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db       8861 ACAGTTAGACACTCCCCTGCCAATTCATGGCTGGGAAACATCATCCAGTACGCCCCGACC 8920

Qy       8581 ATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATGGCTCAAGACACG 8640
             |||||||||||||||||||||||||||||||||||||||||||| |||||||| ||||||||
Db       8921 ATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATCCTCATGGCCCAAGACACT 8980

Qy       8641 CTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTGAGTCCCTTGGAC 8700
             |||||||||||||||| |||||||||| |||||||||||||||||| ||||| ||||||
Db       8981 CTGGACCAGAACCTCAATTTTGAGATGTATGGAGCGGTGTACTCCGTAAGTCCTTTGGAC 9040

Qy       8701 CTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTGCACACATACACT 8760
             |||||||| |||||||||||||||| ||||| |||||||| |||||||||||||| ||
Db       9041 CTCCCAGCCATAATTGAAAGGTTACACGGGCTCGACGCTTTCTCTCTGCACACATACTCT 9100

Qy       8761 CCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGA 8820
             |||||||||||||| |||||||||||||||||||||||||||||||||||| ||||||
Db       9101 CCCCACGAACTGACGCGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCGCCACCTCTCAGA 9160

Qy       8821 GCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGTGGGGGGAGAGCG 8880
             |||||||||| ||||| |||||||||| |||||||||||||||||||||| |||||||||
Db       9161 GCGTGGAAGAGTCGGGCGCGTGCAGTCAGAGCGTCCCTCATCTCCCGTGGAGGGAGAGCG 9220

Qy       8881 GCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 8940
             |||||||||||||| |||||||||||||||||||||| ||||||||||||||||||||||
Db       9221 GCCGTTTGCGGTCGGTATCTCTTCAATTGGGCGGTGAAAACCAAGCTCAAACTCACTCCA 9280

Qy       8941 TTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGCGCCGGCGGGGGC 9000
             |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9281 TTGCCGGAGGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGCGCCGGCGGGGGC 9340

Qy       9001 GACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTTGGCCTACTCCTA 9060
             |||||||||||||||||||||||||||||||||||||||||||| ||| || ||||||||||
Db       9341 GACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTACTCCTTAGCCTACTCCTA 9400

Qy       9061 CTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCGG 9099
             ||||  |||||||||||||| ||||||||||||||||||
Db       9401 CTTTCCGTAGGGGTAGGCCTCTTCCTACTCCCCGCTCGG 9439
```

RESULT 10
AF238481

```
LOCUS       AF238481               9416 bp    RNA     linear   VRL 26-MAR-2000
DEFINITION  Hepatitis C virus 2a polyprotein gene, complete cds.
ACCESSION   AF238481
VERSION     AF238481.1  GI:7329200
KEYWORDS    .
SOURCE      Hepatitis C virus
  ORGANISM  Hepatitis C virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepacivirus.
REFERENCE   1  (bases 1 to 9416)
  AUTHORS   Itakura,J., Nagayama,K., Enomoto,N., Kurosaki,M., Watanabe,H. and
            Sato,C.
  TITLE     Full length cDNA sequence of HCV genotype 2a
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 9416)
  AUTHORS   Itakura,J., Nagayama,K., Enomoto,N., Kurosaki,M., Watanabe,H. and
            Sato,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-FEB-2000) Second Department of Internal Medicine,
            Tokyo Medical and Dental University, 1-5-45 Yushima Bunkyo-Ku,
            Tokyo 113-8519, Japan
FEATURES             Location/Qualifiers
     source          1. .9416
                     /organism="Hepatitis C virus"
                     /mol_type="genomic RNA"
                     /strain="MD2a-1"
                     /db_xref="taxon:11103"
                     /note="genotype: 2a"
     CDS             307. .9408
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="AAF59940.1"
                     /db_xref="GI:7329201"
                     /translation="MSTNPKPQRKTKRNTNRRPQNVKFPGGGQIVGGVYLLPRRGPRL
                     GVRATRKTSERSQPRGRRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRG
                     SRPSWGPNDPRHRSRNVGKVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLED
                     GVNFATGNLPGCSFSIFLLALLSCITIPVSAVQVKNTSNSYMVTNDCSNDSITWQVQD
                     AVLHVPGCVPCEKVENTSRCWIPVSPNVAVRQRGALTQNLRTHIDMVVMSATLCSALY
                     VGDLCGGVMLAAQMFIVSPERHWFVQECNCSIYPGTITGHRMAWDMMMNWSPTTTMIL
```

```
                     AYAMRIPEVIIDIMSGAHWGVMFGLAYFSMQGAWAKVVVILLLTAGVDAETQTVGGAT
                     GRSVLGFTSLFTRGPQQQIQLINTNGSWHINRTALNCNDSLNTGFIASLFYYNNFNSS
                     GCPQRMAACRTIEAFRIGWGTLQYEDNVTNPEDMRPYCWHYPPKQCGIVPARSVCGPV
                     YCFTPSPVVVGTTDKLGVPTYTWGENETDVFLLNSTRPPRGSWFGCTWMNSTGFTKTC
                     GAPPCRTRADFNASTDLLCPTDCFRKHPEATYIKCGSGPWLTPRCLVDYPYRLWHYPC
                     TINYTVFKVRMYVGGVEHRLTAACNFTRGDRCDLEDRDRSQQTPLLHSTTEWAILPCS
                     YSELPALSTGLLHLHQNIVDVQYMYGLSPALTKYVVRWEWVVLLFLLLADARVCACVW
                     MLILLGQAEAALEKLVVLHAASAASCNGFLYFVIFFVAAWYIKGRAVPLAAYSLTGLW
                     PFCLLVLALPQQAYAYDASVHGQXGAVLLILITLFTLTPGYKTLLSRCLWWLCYLLTL
                     GEAMVQEWVPPMQARGGRDGIIWAATIFCPGVVFDITKWLLAVLGPGYLLRGALTRVP
                     YFVRAHALLRMCTVVRHLAGGRYVQMALLALGRWTGTYIYDHLTPMSDWAANGLRDLA
                     VAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREILLGPADGYTSKGWRLL
                     APITAYAQQTRGLLGAIVVSMTGRDKTEQAGEIQILSTVTQSFLGTSIAGVLWTVYHG
                     AGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRNADVIPA
                     RRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGIFRAAVCSRGVAKSIDFIPVET
                     LDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNPSV
                     AATLGFGAYLSKAHGINPNIRTGVRTVTTGEPITYSTYGKFLADGGCAGGAYDIIICD
                     ECHAVDATTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGSEGEIP
                     FYGKAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNAVAYYRGLDVSIIPTQGDV
                     VVVATDALMTGYTGDFDTVIDCNVAVTQVVDFSLDPTFTITTQTVPQDAVSRSQRRGR
                     TGRGRLGIYRYVSTGERASGMFDSVVLGECYDAGAAWYELTPSETTVRLRAYFNTPGL
                     PVCQDHLEFWEAVFTGLTHIDAHFLSQTKQSGENFAYLVAYQATVCARAKAPPPSWDV
                     MWKCLIRLKPTLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAG
                     GVLAAVAAYCLATGCVSIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQ
                     RIAEMLKSKIQGLLQQASKQAHDIQPAVQASWPKVEQFWAKHMWNFISGIQYLAGLST
                     LPGNPAVASMMAFSAALTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAA
                     VGSIGLGKVLVDILAGYGAGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGV
                     ICAAILRRHVGPGEGAVQWMNRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTIT
                     SLLRRLHNWITEDCPIPCAGSWLRDVDWVCTILTDFKNWLTSKLFPKMPGLPFISCQ
                     KGYRGVWAGTGIMTTRCPCGANISGNVRLGSMRITGPKTCMNTWQGTFPINCYTEGQC
                     VPKPAPNFKTAIWRVAASEYAEVTQHGSYSYITGLTTDNVKVPCQLPSPEFFSWVDGV
                     QIHRFAPTPKPFFRDEVSFSVGLNSPVVGSQLPCDPEPDTDVLMSMLTDPSHITAEAA
                     ARRLARGSPPSEASSSASQLSAPSLRATCTTHGRAHDVDMVDANLFMGGDVTRIESES
                     KVVVLDSLDPMAEERSDHEPSIPSEYLLPKSRFPPALPAWARPEYNPPLVEPWKRADY
                     QPPAVAGCALPPPKKTPTPPPRKPRTVGLNESTIADALQQLAIKTFGQPPPSGDSGLS
                     TGADAADSGGQTPPDELAHSETGSISSMPPLEGEPGDPDLESGQVELESPPQGGEVAP
                     GSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEEEKLPINPLSNSLLRYHNKVY
                     CTTAKSASLRAKKVTFDRMQVLDAHYHSVLKDIKLAASKVSARLLTLEEACQLTPPHS
                     ARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSQTPIPTTIMAKNEVFCVDPTKGGK
                     KAARLIVYPDLGVRVCEKMALYDVTQKLPQAVMGASYGFQYSPAQRVEFLLKAWAEKK
                     DPMGFSYDTRCFDSTVTERDIRTEESIYLACSLPEEARTAIHSLTERLYVGGPMFNSK
                     GQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIVAPTMLVCGDDLVVISESQG
                     TEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQGRRRYYLTR
                     DPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWVRMVLMTHFFSILMAQDTLDQNLN
                     FEMYGSVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLRTWKS
                     RARAVRASLISHGGRAAICGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGGDI
                     YHSVSRARPRLLLFSLLLLSVGVGLFLLPAR"
ORIGIN

  Query Match          86.8%;  Score 7895.4;  DB 14;  Length 9416;
  Best Local Similarity  91.7%;  Pred. No. 0;
  Matches 8346;  Conservative    0;  Mismatches  752;  Indels    0;  Gaps    0;

Qy          1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        307 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACTAACCGTCGCCCACAA 366

Qy         61 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        367 AACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 426

Qy        121 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        427 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGG 486

Qy        181 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        487 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAGTCCTGGGGAAAACCAGGA 546

Qy        241 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        547 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 606

Qy        301 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        607 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 666

Qy        361 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        667 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCCGTCGTA 726

Qy        421 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 480
              |||||||||| || ||||||||||||||||||||| ||||||||||||||||||||||
Db        727 GGCGCCCCGCTTGGTGGCGTCGCCAGAGCTCTCGCGCACGGCGTGAGAGTCCTGGAGGAC 786

Qy        481 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 540
              |||||||||||||||||||||||||||||||| |||||||||||||||| ||||||||||
Db        787 GGGGTTAATTTTGCAACAGGGAACTTACCTGGTTGCTCCTTTTCTATTTTCTTGCTGGCC 846

Qy        541 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 600
              || ||||||||||||||||  |||||||| |||| | |||||||||||| ||||| | ||
Db        847 CTACTGTCCTGCATCACCATTCCGGTCTCTGCTGTCCAAGTGAAGAACACCAGTAACAGC 906

Qy        601 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 660
              || |||||||||| || ||  |||| ||||||||||||||||| | ||||  || |||
Db        907 TATATGGTGACTAATGATTGTTCCAACGACAGCATTACCTGGCAGGTTCAGGACGCGGTC 966

Qy        661 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 720
              |||||||||||||||||||||||||||||||||||| ||| |||| | ||||||||||
Db        967 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGAAAATACATCACGGTGCTGGATA 1026

Qy        721 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 780
              |||||||||| || ||||| |||| || || ||||||||||||||  ||||||||||
Db       1027 CCGGTCTCACCAAACGTGGCTGTGCGGCAGCGCGGCGCCCTCACGCAAAACTTGCGGACG 1086

Qy        781 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 840
              |||||||| |||||||||||||||||||||||||| || || ||||||||||||| |||
Db       1087 CACATCGATATGGTTGTGATGTCCGCCACGCTCTGTTCCGCACTCTACGTGGGGGATCTC 1146

Qy        841 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 900
              ||||| |||||||||||||||||||||||||||||||||||||||| ||| ||||||||||
Db       1147 TGCGGCGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGGAGCGCCACTGGTTT 1206

Qy        901 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 960
              || ||||| |||||||||||||||||||| |||||||||||||||| |||||||||||
Db       1207 GTGCAAGAATGCAATTGCTCCATCTACCCTGGCACCATCACTGGACACCGTATGGCATGG 1266

Qy        961 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1020
              || |||||||||||||||||||||||| | |||||||||||||||||||||||||  |
Db       1267 GATATGATGATGAACTGGTCGCCCACGACCACCATGATCTTGGCGTACGCGATGCGCATT 1326

Qy       1021 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1080
              |||||||||||  |||||||||||||||||||||||||||||||||| | |||
Db       1327 CCCGAGGTCATCATAGACATCATGAGCGGGGCTCACTGGGGCGTCATGTTCGGCCTAGCC 1386

Qy       1081 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1140
              |||||||| |||||||||||||||||| ||||||||||||||||| || |||| |||
Db       1387 TACTTCTCCATGCAGGGAGCGTGGGCGAAGGTCGTTGTCATCCTTCTGCTGACCGCTGGG 1446

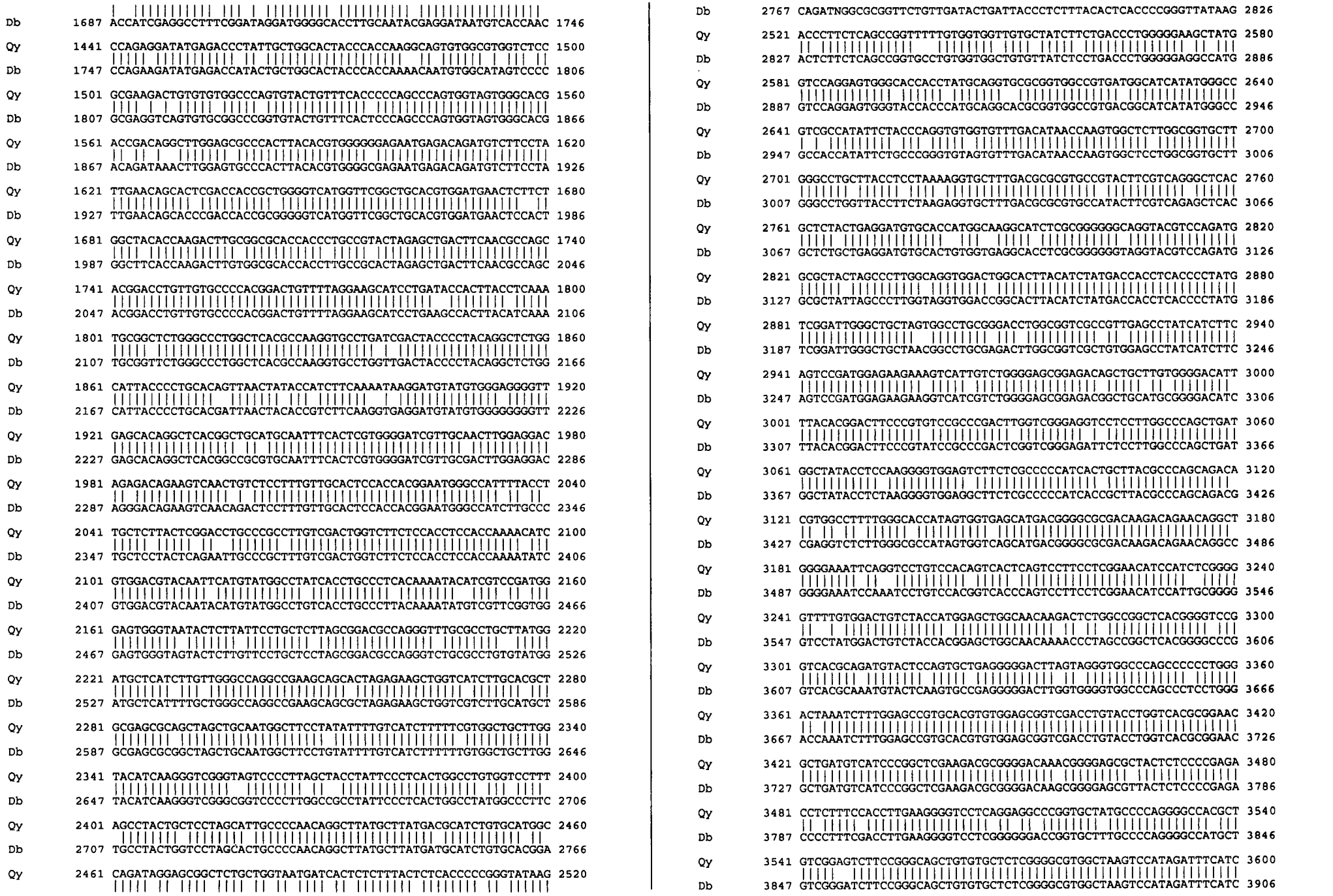
Qy       1141 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1200
              |||||||||   ||||| || |||||||||  | ||| ||  |    | | || |||||
Db       1447 GTGGACGCGGAAACCCAAACAGTTGGGGGTGCAACCGGGCGCAGTGTCTTGGGCTTCACC 1506

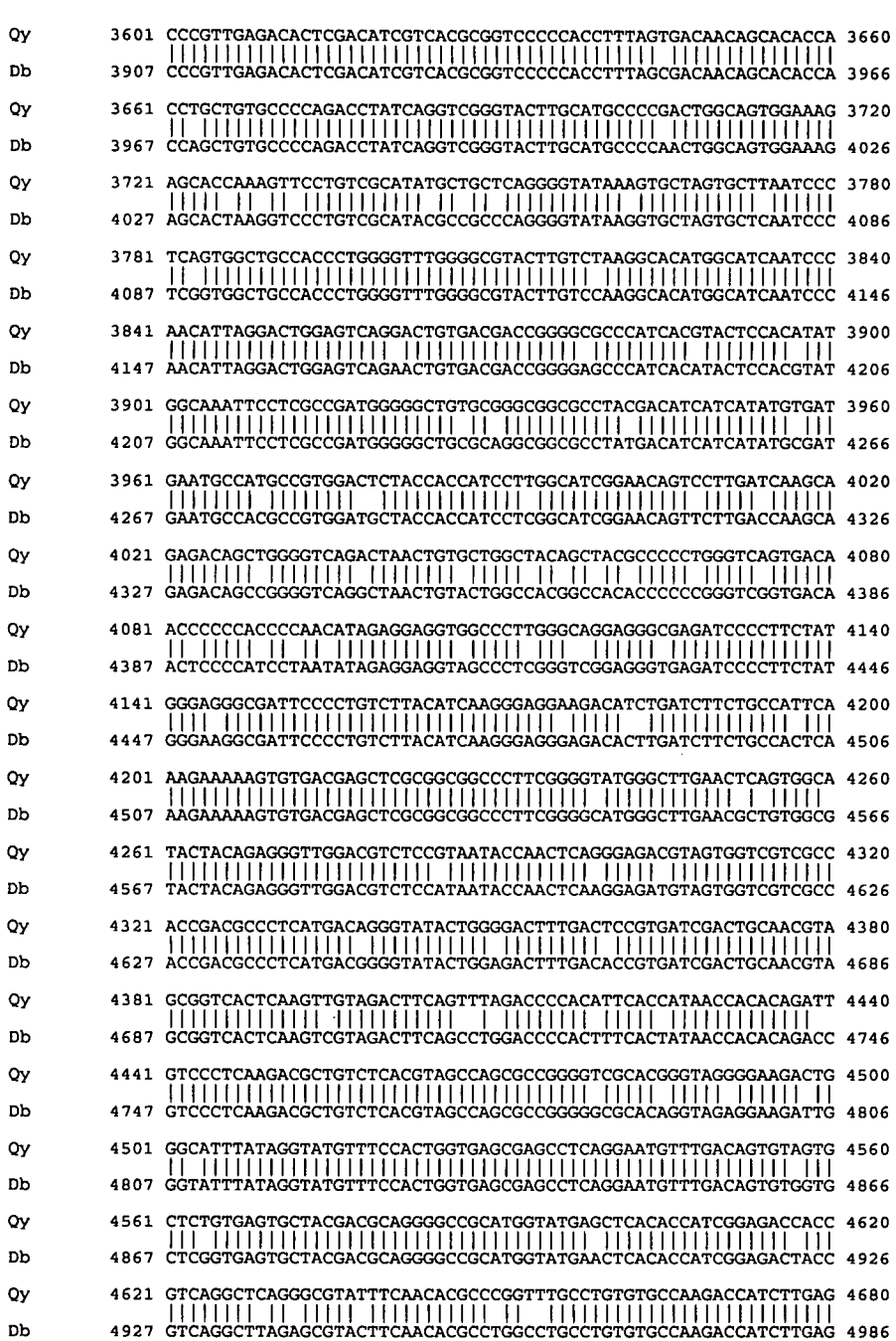
Qy       1201 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1260
              ||| | ||        |||||     ||| | || |||||| | ||||||||||| |||
Db       1507 AGCCTCTTCACGCGTGGCCCTCAACAGCAGATTCAGCTCATCAACACCAATGGCAGTTGG 1566

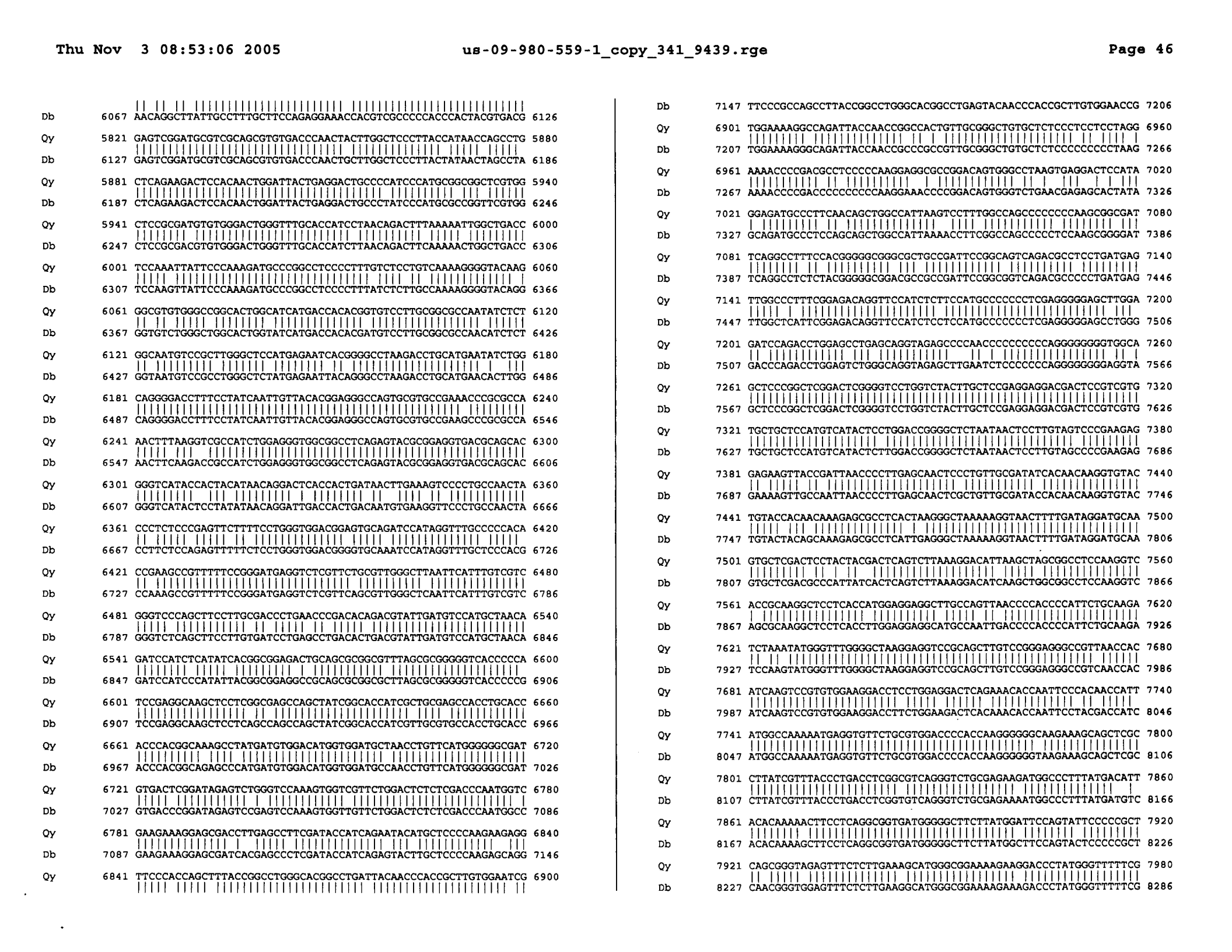
Qy       1261 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1320
              |||||||||||||| |||||||| |||||||||||||||| ||||||||||||||||||
Db       1567 CACATCAACCGCACTGCCCTGAATTGCAATGACTCCTTGAACACCGGCTTTATCGCGTCC 1626

Qy       1321 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1380
              |||||||||  | ||| ||| |||||||||| |||||| |||||||| ||||||| |||
Db       1627 CTGTTCTACTACAACAACTTTAACTCGTCAGGGTGTCCCCAACGCATGGCCGCCTGTCGC 1686

Qy       1381 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1440
```




Qy 3601 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 3660
Db 3907 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGCGACAACAGCACACCA 3966
Qy 3661 CCTGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 3720
Db 3967 CCAGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCAACTGGCAGTGGAAAG 4026
Qy 3721 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 3780
Db 4027 AGCACTAAGGTCCCTGTCGCATACGCCGCCCAGGGGTATAAGGTGCTAGTGCTCAATCCC 4086
Qy 3781 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 3840
Db 4087 TCGGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCCAAGGCACATGGCATCAATCCC 4146
Qy 3841 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 3900
Db 4147 AACATTAGGACTGGAGTCAGAACTGTGACGACCGGGGAGCCCATCACATACTCCACGTAT 4206
Qy 3901 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 3960
Db 4207 GGCAAATTCCTCGCCGATGGGGGCTGCGCAGGCGGCGCCTATGACATCATCATATGCGAT 4266
Qy 3961 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4020
Db 4267 GAATGCCACGCCGTGGATGCTACCACCATCCTCGGCATCGGAACAGTTCTTGACCAAGCA 4326
Qy 4021 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4080
Db 4327 GAGACAGCCGGGGTCAGGCTAACTGTACTGGCCACGGCCACACCCCCCGGGTCGGTGACA 4386
Qy 4081 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4140
Db 4387 ACTCCCCATCCTAATATAGAGGAGGTAGCCCTCGGGTCGGAGGGTGAGATCCCCTTCTAT 4446
Qy 4141 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4200
Db 4447 GGGAAGGCGATTCCCCTGTCTTACATCAAGGGAGGGAGACACTTGATCTTCTGCCACTCA 4506
Qy 4201 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4260
Db 4507 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGCATGGGCTTGAACGCTGTGGCG 4566
Qy 4261 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4320
Db 4567 TACTACAGAGGGTTGGACGTCTCCATAATACCAACTCAAGGAGATGTAGTGGTCGTCGCC 4626
Qy 4321 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4380
Db 4627 ACCGACGCCCTCATGACGGGGTATACTGGAGACTTTGACACCGTGATCGACTGCAACGTA 4686
Qy 4381 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4440
Db 4687 GCGGTCACTCAAGTCGTAGACTTCAGCCTGGACCCCACTTTCACTATAACCACACAGACC 4746
Qy 4441 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4500
Db 4747 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGGCGCACAGGTAGAGGAAGATTG 4806
Qy 4501 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4560
Db 4807 GGTATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTGGTG 4866
Qy 4561 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4620
Db 4867 CTCGGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAACTCACACCATCGGAGACTACC 4926
Qy 4621 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 4680
Db 4927 GTCAGGCTTAGAGCGTACTTCAACACGCCTGGCCTGCCTGTGTGCCAAGACCATCTTGAG 4986
Qy 4681 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 4740
Db 4987 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGACGCCCACTTCCTCTCCCAGACA 5046
Qy 4741 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 4800
Db 5047 AAACAATCGGGGGAGAATTTCGCATACTTAGTAGCCTATCAGGCCACAGTGTGCGCCAGA 5106
Qy 4801 GCCAAAGCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 4860
Db 5107 GCCAAAGCTCCTCCCCCGTCCTGGGACGTTATGTGGAAGTGCTTGATTCGACTCAAGCCC 5166
Qy 4861 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 4920
Db 5167 ACGCTCGTGGGCCCTACACCCCTCCTGTACCGTTTGGGCTCTGTTACCAACGAGGTCACC 5226
Qy 4921 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 4980
Db 5227 CTCACACATCCTGTGACAAAATACATCGCCACATGCATGCAAGCTGACCTTGAGGTCATG 5286
Qy 4981 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5040
Db 5287 ACCAGCACGTGGGTCCTAGCTGGGGGAGTCTTGGCAGCCGTCGCCGCGTATTGCTTGGCG 5346
Qy 5041 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5100
Db 5347 ACCGGGTGTGTTTCCATCATCGGCCGTTTGCACATCAACCAGCGAGCCGTCGTTGCACCG 5406
Qy 5101 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5160
Db 5407 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGCGCCTCTAGAGCGGCT 5466
Qy 5161 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5220
Db 5467 CTCATTGAAGAGGGGCAGCGAATAGCCGAGATGCTGAAATCCAAGATCCAGGGCTTATTG 5526
Qy 5221 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5280
Db 5527 CAGCAAGCTTCTAAACAGGCACATGATATACAACCCGCTGTGCAGGCTTCATGGCCCAAG 5586
Qy 5281 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5340
Db 5587 GTAGAGCAATTCTGGGCCAAACACATGTGGAACTTCATCAGCGGCATTCAATACCTCGCA 5646
Qy 5341 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5400
Db 5647 GGACTATCAACACTGCCAGGGAACCCTGCTGTAGCTTCCATGATGGCATTCAGTGCCGCC 5706
Qy 5401 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5460
Db 5707 CTCACCAGTCCGTTGTCAACTAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5766
Qy 5461 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5520
Db 5767 GCGTCCCAAATCGCGCCACCCGCAGGGGCCACCGGCTTCGTCGTCAGTGGCCTGGTGGGG 5826
Qy 5521 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5580
Db 5827 GCAGCTGTAGGCAGCATAGGCTTGGGTAAGGTGCTGGTGGACATCCTGGCAGGGTATGGT 5886
Qy 5581 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5640
Db 5887 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5946
Qy 5641 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 5700
Db 5947 GAGGATGTCGTCAACCTGCTGCCTGGGATTCTGTCTCCGGGTGCCCTGGTGGTGGGAGTC 6006
Qy 5701 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 5760
Db 6007 ATCTGCGCGGCCATCCTGCGCCGCCACGTGGGGCCGGGGGAAGGCGCGGTTCAGTGGATG 6066
Qy 5761 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 5820



```
Qy       7981 TATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGGACTGAGGAGTCC 8040
             |||||||||||||||| |||||||||||||||||||||| ||||| |||||||| |||
Db       8287 TATGATACCCGATGCTTCGACTCAACCGTCACTGAGAGAGATATCAGAACTGAGGAATCC 8346

Qy       8041 ATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACACTCGCTAACTGAG 8100
             ||||||| |||||||||| |||||||||||||||| |||||||||||| ||||| ||||||
Db       8347 ATATATCTGGCCTGCTCCCTGCCCGAGGAGGCCCGCACTGCCATACATTCGCTGACTGAG 8406

Qy       8101 AGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGCGGGTACAGGCGT 8160
             |||||||||||||||||| |||||||||||||||||||| |||||||||||||||||||
Db       8407 AGACTTTACGTGGGAGGGCCCATGTTCAACAGCAAGGGCCAGACCTGCGGGTACAGGCGT 8466

Qy       8161 TGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACATGCTACGTGAAA 8220
             || |||||||||||||||||| |||||||||| |||||||| |||||||| || |||
Db       8467 TGTCGCGCCAGCGGGGTGCTCACTACTAGCATGGGTAACACCATAACATGCTATGTAAAA 8526

Qy       8221 GCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTGGTATGCGGCGAT 8280
             ||| |||||||||| || || || |||||| | || |||||||||||||| ||||||||
Db       8527 GCCCTAGCGGCTTGCAAGGCCGCGGGGATAGTTGCACCCACAATGCTGGTGTGCGGCGAC 8586

Qy       8281 GACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGGAAACCTGAGAGCC 8340
             |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       8587 GACTTGGTTGTCATCTCAGAAAGCCAGGGGACTGAGGAGGACGAGCGGAACCTGAGAGCC 8646

Qy       8341 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCCAGACCGGAGTAT 8400
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db       8647 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCCAGACCGGAATAT 8706

Qy       8401 GATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTGGGCCCACAAGGC 8460
             || |||||||||||||||||||||||||||||||||||||||||| || |||||||| |||
Db       8707 GACCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCACTCGGCCCACAGGGC 8766

Qy       8461 CGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGGGCTGCCTGGGAA 8520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8767 CGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGGGCTGCCTGGGAA 8826

Qy       8521 ACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAGTACGCCCCGACC 8580
             |||||||||||||||||||| |||||||||||||||||||||||||||| || |||
Db       8827 ACAGTTAGACACTCCCCTGTCAACTCATGGCTGGGAAACATCATCCAGTACGCTCCAACC 8886

Qy       8581 ATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATGGCTCAAGACACG 8640
             ||||||| ||| |||||||||||||||| |||||||||||||||||||| ||||||||
Db       8887 ATATGGGTTCGAATGGTCCTGATGACACATTTCTTCTCCATTCTCATGGCCCAAGACACT 8946

Qy       8641 CTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTGAGTCCCTTGGAC 8700
             |||||||||||||| |||||||| |||||||||| ||||||| |||||||| || ||||||
Db       8947 CTGGACCAGAACCTTAACTTTGAAATGTACGGATCGGTGTATTCCGTGAGCCCTTTGGAC 9006

Qy       8701 CTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTGCACACATACACT 8760
             |||||||| |||||||||||| |||| |||||||||| || ||||||||||||||||||
Db       9007 CTCCCAGCCATAATTGAAAGGCTACACGGGCTTGACGCCTTCTCTCTGCACACATACACT 9066

Qy       8761 CCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGA 8820
             |||||||||||| || ||||||||||||||||||||||||||||||||||||||||||||
Db       9067 CCCCACGAACTAACGCGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGG 9126

Qy       8821 GCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGTGGGGGGAGAGCG 8880
              |||||||||| ||||| |||||||| |||||||||||||||||| ||||||||||||||
Db       9127 ACGTGGAAGAGTCGGGCGCGTGCAGTTAGGGCGTCCCTCATCTCCCATGGGGGGAGAGCG 9186

Qy       8881 GCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 8940
             ||| |||||||||| |||||||||| |||||||||| ||||||||||||||||||||||
Db       9187 GCCATTTGCGGTCGGTATCTCTTCAACTGGGCGGTGAAAACCAAGCTCAAACTCACTCCA 9246

Qy       8941 TTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGCGCCGGCGGGGGC 9000
             |||||||| || |||||||||||| ||||||| ||||| ||||||||||||||||||||
Db       9247 TTGCCGGAGGCACGCCTCCTGGATCTATCCAGTTGGTTTACCGTCGGCGCCGGCGGGGGC 9306

Qy       9001 GACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTTGGCCTACTCCTA 9060
             |||||||||||||||||||||||||||||||||||||||| || |||||| ||||||||||
Db       9307 GACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTGTTACTCTTTAGCCTACTCCTA 9366

Qy       9061 CTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCG 9098
                  ||||  |||||||||||||| ||||||||||||||||
Db       9367 CTTTCCGTAGGGGTAGGCCTCTTCCTACTCCCCGCTCG 9404
```

RESULT 11
AF169004

```
LOCUS       AF169004               9653 bp    RNA     linear   VRL 06-SEP-2001
DEFINITION  Hepatitis C virus isolate G2aK3 polyprotein gene, complete cds.
ACCESSION   AF169004
VERSION     AF169004.1  GI:6707283
KEYWORDS    .
SOURCE      Hepatitis C virus
  ORGANISM  Hepatitis C virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepacivirus.
REFERENCE   1  (bases 1 to 9653)
  AUTHORS   Kurihara,C., Ishiyama,N., Nishiyama,Y., Fukushi,S., Kageyama,T.,
            Katayama,K. and Miura,S.
  TITLE     Molecular characterization of hepatitis C virus genotype 2a from
            the entire sequences of four isolates
  JOURNAL   J. Med. Virol. 64 (4), 466-475 (2001)
  MEDLINE   21361470
   PUBMED   11468731
REFERENCE   2  (bases 1 to 9653)
  AUTHORS   Katayama,K. and Kurihara,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-1999) Basic Research Division, BioMedical
            Laboratories, Inc., 1361-1 Matoba, Kawagoe, Saitama 350-1101, Japan
FEATURES             Location/Qualifiers
     source          1. .9653
                     /organism="Hepatitis C virus"
                     /mol_type="genomic RNA"
                     /isolate="G2aK3"
                     /db_xref="taxon:11103"
                     /note="genotype: 2a"
     CDS             341. .9442
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="AAF25612.1"
                     /db_xref="GI:6707284"
                     /translation="MSTKPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRL
                     GVRATRKTSERSQPRGRRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRG
                     SRPSWGPTDPRHRSRNVGKVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLED
                     GVNYATGNLPGCSFSIFLLALLSCITIPVSAAQVKNTSTGYMVTNDCSNDSITWQLQA
                     AVLHVPGCVPCEKVGNXSRCWIPVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALY
                     VGDLCGGVMLAAQMFIVSPQHHWFVQECNCSIYPGTITGHRMAWDMMMNWSPTATMIL
                     AYAMRVPEVIIDIIGGAHWGVMFGLAYFSMQGAWAKVVVILLLAAGVDAXTXTXAGXX
                     AXTVSGIXXLFXPGAKQNXQLINSNGSWHINRTALNCNDSLNTGFLASLFYTHSFNSS
                     GCPERLSACRNIEAFRIGWGTLQYEENVTNPEDMRPYCWHYPPKQCGIVPAESVCGPV
                     YCFTPSPVVVGTTDRLGVPTYTWGENETDVFLLNSTRPPSGSWFGCTWMNSTGFTKAC
                     GAPPCRXRADFNASTDLLCPTDCFRKHPEATYMKCGSGPWLTPRCLVDYPYRLWHYPC
                     TVNYTIFKIRMYVGGVEHRLTAACNFSRGYRCNLEDRDRGQVSPLLHSTTEWAILPCS
                     YTDLPALSTGLIHLHQNIVDVQYMYGLSPALTKYVVRWEWVVLLFLLLADARVCACLW
                     MLILLGQAEAALEKLVVLHAASAASCNGFLYFVIFFIAAWYIKGRXVPLATYSLTGLW
                     PFCLLLLALPQQAYAYDGSVHGQIGAALLILITLFTLTPAYKTLLSRCLWWLCYLLTL
                     GEAMVQEWAPSMQARGGRDGIIWAATIFCPGVVFDITKWLLAVLGPGYLLRDALTRVP
                     YFVRAHALLRMCTXVRHLAGGRYVQMALLALGRWTGTYIYDHLTPMSDWAASGLRDLA
                     VAVEPIIFSPMEKKVITWGAETAACGDILHGLPVSARLGREILLGPADGYTSKGWRLL
                     APITAYAQQTRGLLGAIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSIXGVLWTVYHG
                     AGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRNADVIPA
                     RRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGIFRAAVCSRGVAKSIDFIPVEA
                     LAIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNPSV
                     AATLGFGAYLSRAHGINPNIRTGVRTVTTGAAITYSTYGKFLADGGCXGGAYDIIICD
                     ECHSVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIP
                     FYGKAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNAVAYYRGLDVSVIPTQGDV
                     VVVATDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQTVPQDAVSRSQRRGR
                     TGRGRLGIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPVETTVRLRAYFNTPGL
                     PVCQDHLEFWEAVFTGLTHIDAHFLSQTKQSGENFAYLVAYQATVCARAKAPPPSWDI
                     MWKCLTRLKPTLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAG
                     GVLAAVAAYCLATGCVSIIGRLHXNQQVVVAPDKEVLYEAFDEMEECASRAALLEEGQ
                     RIAEMLKSKIQGLLQQASKQAQDIQPAVQASWPRVEQFWAKHMWNFISGIQYLAGLST
                     LPGNPAVASMMAFSAALTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAA
```

```
                     VGSIGLGKVLVDILAGYGAGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGV
                     ICAAILRRHVGPGEGAVQWMNRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTIT
                     SLLRRLHNWITEDCPIPCAGSWLRDVWDWVCSILIDFKNWLTSKLFPKMPGLPLISCQ
                     KGYKGVWAGTGIMTTRCPCGANISGNVRLGSMRITGPKTCMNTWHGTFPINCYTEGQC
                     VPKPAPNFKTAIWRVAASEYAEVTQHGSYSYITGLTTDNLKVPCQLPSPEFFSWVDGV
                     QIHRFAPTPKPFFRDEVSFCVGLNSFVVGSQLPCDPEPDTDVLTSMLTDPSHITAEAA
                     ARRLARGSPPSEASSSASQLSAPSLRATCTTHGKTYDVDMVDANLFMGGDVTRIESES
                     KVIVLDSLDSMVEEGSDHEPSIPSEYMLPRTKFPPALPVWARPDYNPPVVESWKRPDY
                     QPPTVAGCALPPSKETPTPPPRRRRTVGLSESTIGDALQQLAIKTFGQPPSTGDSGXS
                     TGADAADSGGRTPPDELTPSETGSTSSMPPLEGEPGDPDLEPEQVELQPPPQEGXVAP
                     GSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEEEKLPINPXSNSLLRYHNKVY
                     CTTSKSASLRAKKVTFDRMQVLDAHYDSVLKDIKLAASKVSARLLTVEEACQLTPPHS
                     AKSKYGFGAKEVRSLSRRAVNHIKSVWKDLLEDSGTPIPTTIMAKNEVFCVEPTKGGK
                     KAARLIVYPDLGVRVCEKMALYDVTQKLPQAVMGASYGFQYSPAQRVEFLLKAWADKK
                     DPMGFSYDTRCFDSTVTERDIRTEESIYQACSLPEEARTAIHSLTERLYVGGPMFNSK
                     GQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIVAPTMLVCGDDLVVISESQG
                     AEEDERNLRVFTEAMTRYSAPSGDPPRPEYDLELITSCSSNVSVARDPQGRRRYYLTR
                     DPTTPIARAAWETVRHFPVNSWLGNIIQYAPTIWVRMVLMTHFFSILMAQDTLDQNLN
                     FEMYGSVYSVSPSDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLRAWKS
                     RARAVRASLISHGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGGDI
                     YLCVLRARPRLLLLSLLLLSVGLGLFLLPAR"
ORIGIN

  Query Match           86.8%;  Score 7893.6;  DB 14;  Length 9653;
  Best Local Similarity   91.4%;  Pred. No. 0;
  Matches 8317;  Conservative   39;  Mismatches  742;  Indels    0;  Gaps    0;

Qy          1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 60
              |||||||||| |||||||||||||||||||||||||||| || ||||||||||||||||
Db        341 ATGAGCACAAAACCTAAACCTCAAAGAAAAACCAAAAGAAATACTAACCGTCGCCCACAA 400

Qy         61 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 460

Qy        121 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGG 520

Qy        181 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 240
              |||||||||||||||| |||||||||||||||||||||| ||||||||||||||||||||
Db        521 AGGCGCCAGCCCATCCCCAAAGATCGGCGCTCCACTGGCAAGTCCTGGGGAAAACCAGGA 580

Qy        241 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 300
              |||||||||||||||| |||||||||||||||||||||||| ||||||||||||||||||
Db        581 TACCCCTGGCCCCTATATGGGAATGAGGGACTCGGCTGGGCAGGGTGGCTCCTGTCCCCC 640

Qy        301 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 360
              |||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        641 CGAGGYTCCCGTCCCTCTTGGGGCCCCACTGACCCCCGGCATAGGTCGCGCAACGTGGGT 700

Qy        361 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 420
              |||||||||||||||||||||||||||||||||||||||||||||||| || |||||||
Db        701 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATTCCCGTCGTA 760

Qy        421 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 480
              |||||||||| || |||||||||||||||||||||||| |||||||||||||||||||||
Db        761 GGCGCCCCGCTTGGTGGCGTCGCCAGAGCTCTCGCGCATGGTGTGAGAGTCCTGGAGGAC 820

Qy        481 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 540
              |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db        821 GGGGTTAATTATGCAACAGGGAACTTACCTGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 880

Qy        541 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 600
              || |||||||||||||| ||||||| |||||| | |||||||||| ||| |||||||||
Db        881 CTACTGTCCTGCATCACCATCCCGGTCTCTGCTGCCCAGGTGAAGAACACCAGCACCGGT 940

Qy        601 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 660
              |||||||||| || |||||  |||||||||||| |||||||||| |||||||| |||||
Db        941 TACATGGTGACCAATGACTGTTCCAATGACAGCATCACCTGGCAGCTTCAGGCTGCGGTC 1000

Qy        661 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 720
              |||||||||||||||| |||||||| |||||||||| ||| :||| | ||| ||||||

Db       1001 CTCCACGTCCCCGGGTGTGTCCCGTGTGAGAAAGTGGGAAATAYATCACGGTGTTGGATA 1060

Qy        721 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 780
              |||||||||| || ||||| |||||||||||| |||||||||||||||||||||||||||
Db       1061 CCGGTCTCACCAAACGTGGCTGTGCAGCGGCCAGGCGCCCTCACGCAGGGCTTGCGGACG 1120

Qy        781 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 840
              |||||||||||||||||||||||||||||||| |||||||| ||||||||||||||||||
Db       1121 CACATCGACATGGTTGTGATGTCCGCCACGCTGTGCTCTGCTCTCTACGTGGGGGACCTC 1180

Qy        841 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 900
              ||||| |||||||||||||||||||| |||||||| ||||||||||||||||||||||||
Db       1181 TGCGGCGGGGTGATGCTCGCAGCCCAGATGTTCATCGTCTCGCCGCAGCACCACTGGTTT 1240

Qy        901 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 960
              || || || |||||||||||||||||||||||||||||:|||||||||| |||||||||
Db       1241 GTGCAGGAATGCAATTGCTCCATCTACCCTGGTACCATYACTGGACACCGTATGGCATGG 1300

Qy        961 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1020
              |||||||||||||||||||||||||||||| |||||||| |||||||||||||||| ||
Db       1301 GACATGATGATGAACTGGTCGCCCACGGCCACCATGATCCTGGCGTACGCGATGCGCGTT 1360

Qy       1021 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1080
              |||||||||| |||||||||||| |||||||| |||||||||||||||||||| |||
Db       1361 CCCGAGGTCATCATAGACATCATTGGCGGGGCTCACTGGGGCGTCATGTTCGGCTTAGCC 1420

Qy       1081 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1140
              |||||||| |||||||||||||||||||||||||||||||||||||||:|| ||
Db       1421 TACTTCTCCATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCYGCTGGA 1480

Qy       1141 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1200
              |||||||||:::|||||: |||:||| |||| ::|:|||||| ||| | |  || ||:||
Db       1481 GTGGACGCGVVVACCCRCACTRTTGCGGGTASWGSCGCGCAMACCGTCAGCGGCATCRCC 1540

Qy       1201 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1260
              ||  | ||        ||: | | |||||| ||:|||||  ||||| |||||||||| |||
Db       1541 AGKCTKTTCWCGCCCGGBGCTAAGCAGAACATSCAGCTTATTAACTCCAATGGCAGTTGG 1600

Qy       1261 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1320
              || |||||||||| |||||||| |||||||||||| || |||||||||| |||||||
Db       1601 CATATCAACCGCACTGCCCTGAATTGCAATGACTCCCTGAACACCGGCTTTCTCGCGTCC 1660

Qy       1321 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1380
              ||||||||||||||||:|||||||||||||||||||||||||| ||||||||||||
Db       1661 CTGTTCTACACCCACAGYTTCAACTCGTCAGGATGTCCCGAACGCCTGTCCGCCTGCCGT 1720

Qy       1381 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1440
              | |||||||||||| ||| | |||||||| |||||||||| ||||| ||||||||||
Db       1721 AATATCGAGGCCTTTCGGATAGGATGGGGCACCTTGCAATACGAGGAGAATGTCACCAAC 1780

Qy       1441 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1500
              ||||| |||||||||| |||||||||:|| |||||||  |||||||  | ||| ||
Db       1781 CCAGAAGATATGAGACCATATTGCTGGCAYTATCCACCAAAACAGTGTGGTATAGTCCCC 1840

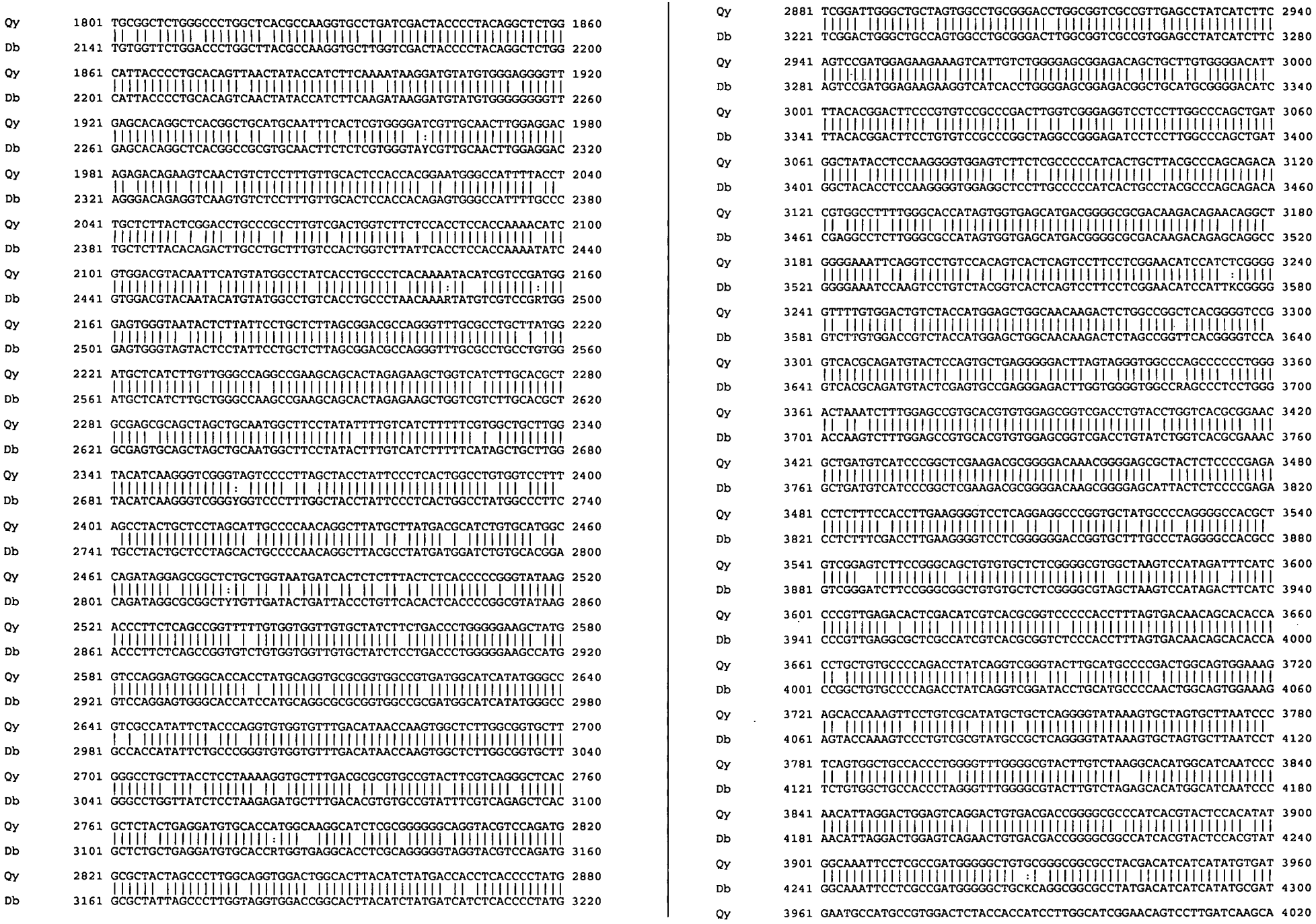
Qy       1501 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1560
              ||| || ||||||| || |||||||||||||||||||||||:||:|||||||||
Db       1841 GCGGAGTCTGTGTGCGGTCCAGTGTACTGTTTCACCCCCAGCCCAGTRGTRGTGGGCACG 1900

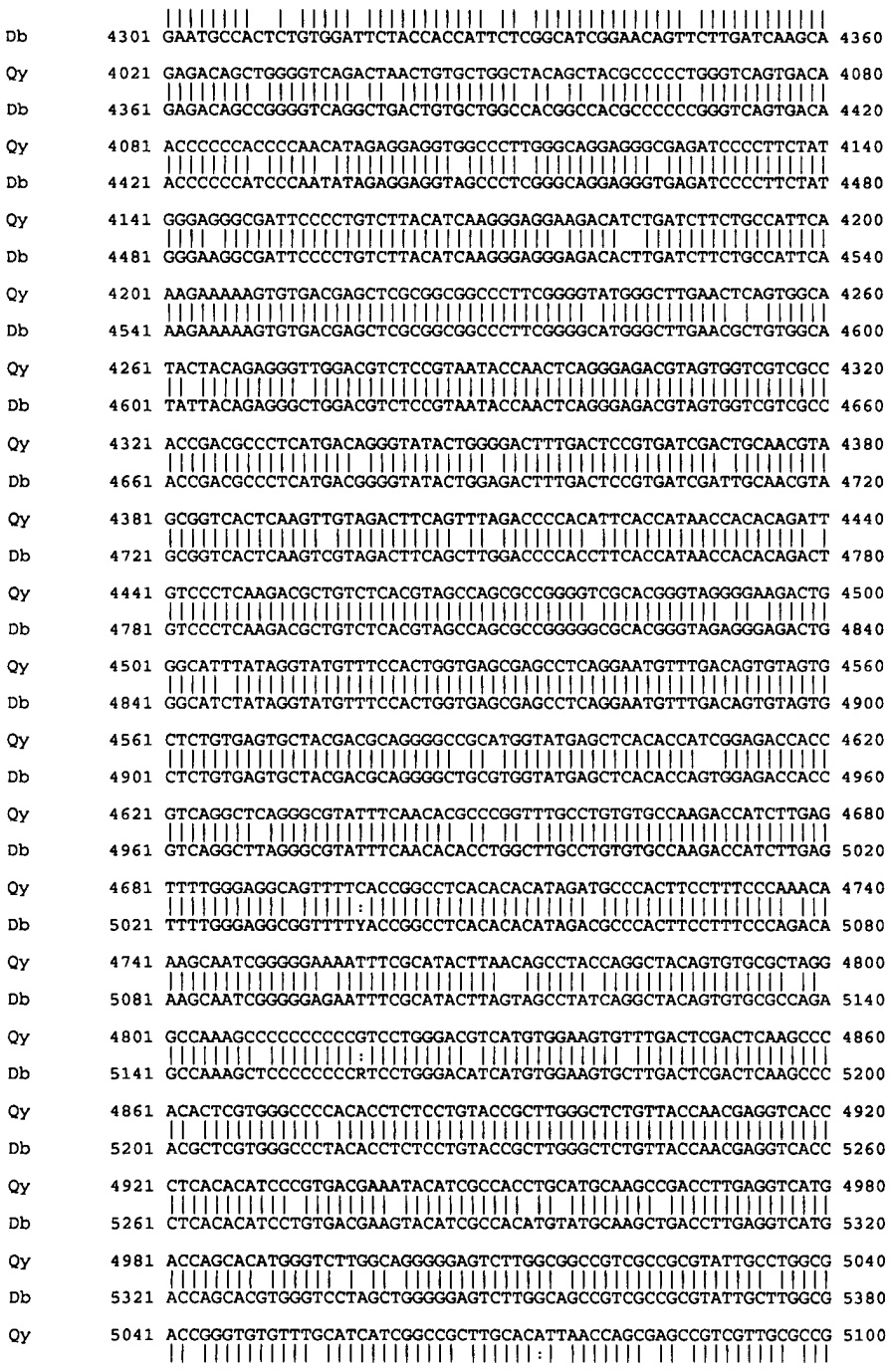
Qy       1561 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1620
              ||||| |||||||||| ||| |||||||||||||| |||||||||||||||||||||||
Db       1901 ACCGATAGGCTTGGAGTGCCTACTTACACGTGGGGAGAGAATGAGACAGATGTCTTCCTA 1960

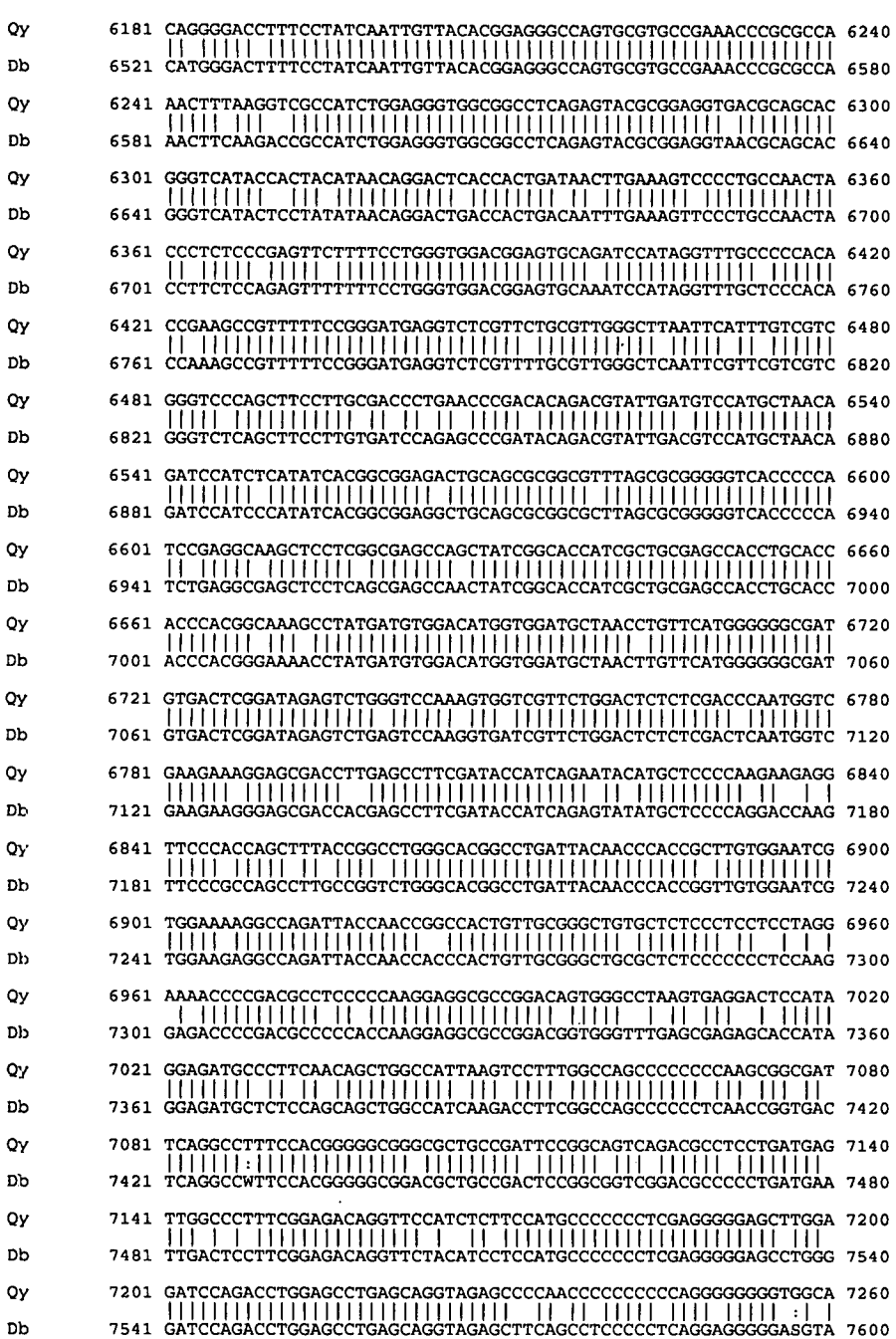
Qy       1621 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 1680
              |||||||||| ||||| ||| |||||||||||||||||||||||||||||||||  ||
Db       1961 TTGAACAGCACCCGACCGCCGTCGGGGTCATGGTTCGGCTGCACGTGGATGAACTCCACT 2020

Qy       1681 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 1740
              || | ||||||| |||| |||||||||| ||||| |:||||||||| ||:|| ||||||
Db       2021 GGTTTCACCAAGGCTTGTGGCGCACCACCTTGCCGCAYTAGAGCTGATTTYAATGCCAGC 2080

Qy       1741 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 1800
              |||||| |||||||||||||| ||||||||||||||||||||||  |||||||| | |||
Db       2081 ACGGACTTGTTGTGCCCCACCGACTGTTTTAGGAAGCATCCTGAAGCCACTTACATGAAA 2140
```




Db 4301 GAATGCCACTCTGTGGATTCTACCACCATTCTCGGCATCGGAACAGTTCTTGATCAAGCA 4360
Qy 4021 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4080
Db 4361 GAGACAGCCGGGGTCAGGCTGACTGTGCTGGCCACGGCCACGCCCCCCGGGTCAGTGACA 4420
Qy 4081 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4140
Db 4421 ACCCCCCATCCCAATATAGAGGAGGTAGCCCTCGGGCAGGAGGGTGAGATCCCCTTCTAT 4480
Qy 4141 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4200
Db 4481 GGGAAGGCGATTCCCCTGTCTTACATCAAGGGAGGGAGACACTTGATCTTCTGCCATTCA 4540
Qy 4201 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4260
Db 4541 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGCATGGGCTTGAACGCTGTGGCA 4600
Qy 4261 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4320
Db 4601 TATTACAGAGGGCTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4660
Qy 4321 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4380
Db 4661 ACCGACGCCCTCATGACGGGGTATACTGGAGACTTTGACTCCGTGATCGATTGCAACGTA 4720
Qy 4381 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4440
Db 4721 GCGGTCACTCAAGTCGTAGACTTCAGCTTGGACCCCACCTTCACCATAACCACACAGACT 4780
Qy 4441 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4500
Db 4781 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGGCGCACGGGTAGAGGGAGACTG 4840
Qy 4501 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4560
Db 4841 GGCATCTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4900
Qy 4561 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4620
Db 4901 CTCTGTGAGTGCTACGACGCAGGGGCTGCGTGGTATGAGCTCACACCAGTGGAGACCACC 4960
Qy 4621 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 4680
Db 4961 GTCAGGCTTAGGGCGTATTTCAACACACCTGGCTTGCCTGTGTGCCAAGACCATCTTGAG 5020
Qy 4681 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 4740
Db 5021 TTTTGGGAGGCGGTTTTYACCGGCCTCACACACATAGACGCCCACTTCCTTTCCCAGACA 5080
Qy 4741 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 4800
Db 5081 AAGCAATCGGGGGAGAATTTCGCATACTTAGTAGCCTATCAGGCTACAGTGTGCGCCAGA 5140
Qy 4801 GCCAAAGCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 4860
Db 5141 GCCAAAGCTCCCCCCCCRTCCTGGGACATCATGTGGAAGTGCTTGACTCGACTCAAGCCC 5200
Qy 4861 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 4920
Db 5201 ACGCTCGTGGGCCCTACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 5260
Qy 4921 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 4980
Db 5261 CTCACACATCCTGTGACGAAGTACATCGCCACATGTATGCAAGCTGACCTTGAGGTCATG 5320
Qy 4981 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5040
Db 5321 ACCAGCACGTGGGTCCTAGCTGGGGGAGTCTTGGCAGCCGTCGCCGCGTATTGCTTGGCG 5380
Qy 5041 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5100
Db 5381 ACTGGGTGTGTTTCCATCATCGGCCGGTTGCACRTCAACCAGCAAGTCGTCGTTGCACCG 5440
Qy 5101 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5160
Db 5441 GACAAGGAGGTCCTTTATGAGGCTTTTGACGAGATGGAGGAATGTGCCTCTAGAGCGGCT 5500
Qy 5161 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5220
Db 5501 CTCCTTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5560
Qy 5221 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5280
Db 5561 CAGCAAGCCTCCAAACAGGCCCAGGACATACAACCCGCTGTGCAGGCTTCATGGCCCAGA 5620
Qy 5281 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5340
Db 5621 GTGGAGCAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATTCAATATCTCGCA 5680
Qy 5341 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5400
Db 5681 GGACTGTCAACRCTCCCAGGGAACCCTGCTGTGGCTTCCATGATGGCGTTCAGTGCCGCC 5740
Qy 5401 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5460
Db 5741 CTTACCAGTCCACTGTCGACTAGCACCACTATCCTTCTCAACATCTTRGGAGGCTGGCTG 5800
Qy 5461 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5520
Db 5801 GCGTCCCAAATTGCGCCGCCCGCGGGGGCCACTGGCTTCGTCGTCAGTGGCCTGGTGGGG 5860
Qy 5521 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5580
Db 5861 GCTGCCGTAGGCAGCATCGGCCTGGGTAAGGTGCTGGTGGATATCCTGGCGGGGTACGGT 5920
Qy 5581 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5640
Db 5921 GCGGGCATTTCGGGGGCCCTCGTCGCATTTAAGATCATGTCTGGCGAGAAGCCCTCCATG 5980
Qy 5641 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 5700
Db 5981 GAGGATGTCGTCAACCTGCTGCCCGGGATTCTGTCTCCGGGTGCCCTGGTGGTGGGAGTC 6040
Qy 5701 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 5760
Db 6041 ATCTGCGCGGCCATTCTGCGCCGCCACGTGGGTCCGGGGGAGGGCGCGGTCCAATGGATG 6100
Qy 5761 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 5820
Db 6101 AACAGGCTCATTGCCTTCGCTTCCAGAGGAAACCACGTCGCCCCTACCCATTACGTGACG 6160
Qy 5821 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 5880
Db 6161 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTGCTTGGCTCTCTTACCATAACTAGCCTA 6220
Qy 5881 CTCAGAAGACTCCACAACTGGATTACTGAGGACTGCCCCATCCCATGCGGCGGCTCGTGG 5940
Db 6221 CTCAGGAGACTCCATAACTGGATTACTGAGGACTGCCCCATTCCATGCGCCGGCTCGTGG 6280
Qy 5941 CTCCGCGATGTGTGGGACTGGGTTTGCACCATCCTAACAGACTTTAAAAATTGGCTGACC 6000
Db 6281 CTCCGCGACGTGTGGGACTGGGTTTGTTCCATCCTGATAGACTTCAAAAATTGGCTGACT 6340
Qy 6001 TCCAAATTATTCCCAAAGATGCCCGGCCTCCCCTTTGTCTCCTGTCAAAAGGGGTACAAG 6060
Db 6341 TCCAAGTTGTTCCCRAAGATGCCCGGCCTCCCCTTAATCTCTTGCCAAAAGGGGTACAAG 6400
Qy 6061 GGCGTGTGGGCCGGCACTGGCATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6120
Db 6401 GGCGTGTGGGCCGGCACTGGTATCATGACCACACGGTGTCCTTGCGGCGCTAACATCTCT 6460
Qy 6121 GGCAATGTCCGCTTGGGCTCCATGAGAATCACGGGGCCTAAGACCTGCATGAATATCTGG 6180
Db 6461 GGCAATGTCCGCCTGGGCTCTATGAGAATCACAGGGCCCAAAACCTGCATGAACACTTGG 6520


Qy 6181 CAGGGGACCTTTCCTATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCGCCA 6240
Db 6521 CATGGGACTTTTCCTATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCGCCA 6580
Qy 6241 AACTTTAAGGTCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6300
Db 6581 AACTTCAAGACCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTAACGCAGCAC 6640
Qy 6301 GGGTCATACCACTACATAACAGGACTCACCACTGATAACTTGAAAGTCCCCTGCCAACTA 6360
Db 6641 GGGTCATACTCCTATATAACAGGACTGACCACTGACAATTTGAAAGTTCCCTGCCAACTA 6700
Qy 6361 CCCTCTCCCGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCACA 6420
Db 6701 CCTTCTCCAGAGTTTTTTTCCTGGGTGGACGGAGTGCAAATCCATAGGTTTGCTCCCACA 6760
Qy 6421 CCGAAGCCGTTTTTCCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6480
Db 6761 CCAAAGCCGTTTTTCCGGGATGAGGTCTCGTTTTGCGTTGGGCTCAATTCGTTCGTCGTC 6820
Qy 6481 GGGTCCCAGCTTCCTTGCGACCCTGAACCCGACACAGACGTATTGATGTCCATGCTAACA 6540
Db 6821 GGGTCTCAGCTTCCTTGTGATCCAGAGCCCGATACAGACGTATTGACGTCCATGCTAACA 6880
Qy 6541 GATCCATCTCATATCACGGCGGAGACTGCAGCGCGGCGTTTAGCGCGGGGGTCACCCCCA 6600
Db 6881 GATCCATCCCATATCACGGCGGAGGCTGCAGCGCGGCGCTTAGCGCGGGGGTCACCCCCA 6940
Qy 6601 TCCGAGGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 6660
Db 6941 TCTGAGGCGAGCTCCTCAGCGAGCCAACTATCGGCACCATCGCTGCGAGCCACCTGCACC 7000
Qy 6661 ACCCACGGCAAAGCCTATGATGTGGACATGGTGGATGCTAACCTGTTCATGGGGGGCGAT 6720
Db 7001 ACCCACGGGAAAACCTATGATGTGGACATGGTGGATGCTAACTTGTTCATGGGGGGCGAT 7060
Qy 6721 GTGACTCGGATAGAGTCTGGGTCCAAAGTGGTCGTTCTGGACTCTCTCGACCCAATGGTC 6780
Db 7061 GTGACTCGGATAGAGTCTGAGTCCAAGGTGATCGTTCTGGACTCTCTCGACTCAATGGTC 7120
Qy 6781 GAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCAGAATACATGCTCCCCAAGAAGAGG 6840
Db 7121 GAAGAAGGGAGCGACCACGAGCCTTCGATACCATCAGAGTATATGCTCCCCAGGACCAAG 7180
Qy 6841 TTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCGCTTGTGGAATCG 6900
Db 7181 TTCCCGCCAGCCTTGCCGGTCTGGGCACGGCCTGATTACAACCCACCGGTTGTGGAATCG 7240
Qy 6901 TGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTCCCTCCTCCTAGG 6960
Db 7241 TGGAAGAGGCCAGATTACCAACCACCCACTGTTGCGGGCTGCGCTCTCCCCCCCTCCAAG 7300
Qy 6961 AAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGTGAGGACTCCATA 7020
Db 7301 GAGACCCCGACGCCCCCACCAAGGAGGCGCCGGACGGTGGGTTTGAGCGAGAGCACCATA 7360
Qy 7021 GGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCCCCAAGCGGCGAT 7080
Db 7361 GGAGATGCTCTCCAGCAGCTGGCCATCAAGACCTTCGGCCAGCCCCCCTCAACCGGTGAC 7420
Qy 7081 TCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACGCCTCCTGATGAG 7140
Db 7421 TCAGGCCWTTCCACGGGGGCGGACGCTGCCGACTCCGGCGGTCGGACGCCCCCTGATGAA 7480
Qy 7141 TTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCCTCGAGGGGGAGCCTGGA 7200
Db 7481 TTGACTCCTTCGGAGACAGGTTCTACATCCTCCATGCCCCCCCCTCGAGGGGGAGCCTGGG 7540
Qy 7201 GATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCCAGGGGGGGGTGGCA 7260
Db 7541 GATCCAGACCTGGAGCCTGAGCAGGTAGAGCTTCAGCCTCCCCCTCAGGAGGGGGASGTA 7600
Qy 7261 GCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGACGACTCCGTCGTG 7320
Db 7601 GCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGATGACTCCGTCGTG 7660
Qy 7321 TGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGTAGTCCCGAAGAG 7380
Db 7661 TGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGTAGCCCCGAGGAG 7720
Qy 7381 GAGAAGTTACCGATTAACCCCTTGAGCAACTCCCTGTTGCGATATCACAACAAGGTGTAC 7440
Db 7721 GAAAAGTTGCCAATTAACCCCWTGAGCAACTCGCTGTTGCGATACCACAACAAGGTGTAC 7780
Qy 7441 TGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTTGATAGGATGCAA 7500
Db 7781 TGTACTACATCAAAGAGCGCCTCACTGAGGGCTAAAAAGGTAACTTTTGATAGGATGCAA 7840
Qy 7501 GTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCGGCCTCCAAGGTC 7560
Db 7841 GTGCTCGACGCCCATTATGACTCAGTYTTAAAGGACATCAAGCTAGCGGCTTCCAAGGTT 7900
Qy 7561 ACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCCCATTCTGCAAGA 7620
Db 7901 AGTGCAAGGCTCCTCACCGTGGAGGAGGCGTGCCAGCTTACTCCACCCCATTCTGCAAAA 7960
Qy 7621 TCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGGGCCGTTAACCAC 7680
Db 7961 TCCAAGTATGGGTTTGGAGCTAAGGAGGTCCGCAGCTTGTCCAGGAGGGCCGTCAACCAC 8020
Qy 7681 ATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATTCCCACAACCATT 7740
Db 8021 ATCAAGTCCGTGTGGAAGGACCTCCTGGAAGACTCAGGAACACCAATTCCTACGACCATA 8080
Qy 7741 ATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAGAAAGCAGCTCGC 7800
Db 8081 ATGGCCAAAAATGAGGTGTTCTGCGTGGAACCCACCAAGGGGGGTAAGAAAGCAGCTCGC 8140
Qy 7801 CTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCCCTTTATGACATT 7860
Db 8141 CTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCCCTTTATGATGTC 8200
Qy 7861 ACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAGTATTCCCCCGCT 7920
Db 8201 ACACAAAAGCTTCCTCAGGCGGTGATGGGGGCTTCTTATGGCTTCCAGTACTCCCCTGCC 8260
Qy 7921 CAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCTATGGGTTTTTCG 7980
Db 8261 CAACGGGTGGAGTTCCTCTTGAAGGCATGGGCGGATAAGAAAGACCCTATGGGTTTTTCG 8320
Qy 7981 TATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGGACTGAGGAGTCC 8040
Db 8321 TATGATACCCGATGCTTTGACTCAACTGTCACTGAGAGGGACATCAGAACTGAGGAGTCC 8380
Qy 8041 ATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACACTCGCTAACTGAG 8100
Db 8381 ATATACCAGGCCTGCTCCCTGCCCGAGGAGGCCCGCACTGCCATACACTCGCTGACTGAG 8440
Qy 8101 AGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGCGGGTACAGGCGT 8160
Db 8441 AGACTTTACGTGGGAGGGCCCATGTTTAACAGCAAGGGTCAGACCTGCGGGTACAGGCGT 8500
Qy 8161 TGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACATGCTACGTGAAA 8220
Db 8501 TGCCGCGCCAGCGGGGTGCTCACTACTAGCATGGGGAACACCATCACATGCTACGTGAAA 8560
Qy 8221 GCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTGGTATGCGGCGAT 8280
Db 8561 GCCCTGGCGGCTTGTAAGGCTGCGGGGATAGTTGCGCCCACAATGCTGGTATGCGGCGAC 8620
Qy 8281 GACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGGAACCTGAGAGCC 8340
Db 8621 GACTTGGTTGTCATCTCGGAAAGCCAGGGGGCTGAGGAGGACGAGCGGAACTTGAGAGTC 8680
Qy 8341 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCCAGACCGGAGTAT 8400

```
Db       8681 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTTCCGGCGACCCCCCCAGACCGGAATAT 8740

Qy       8401 GATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTGGGCCCACAAGGC 8460

Db       8741 GACCTGGAGCTGATAACATCTTGCTCTTCAAATGTGTCTGTGGCACGGGACCCACAGGGC 8800

Qy       8461 CGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGGGCTGCCTGGGAA 8520

Db       8801 CGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGGGCTGCCTGGGAA 8860

Qy       8521 ACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAGTACGCCCCGACC 8580

Db       8861 ACAGTCAGACACTTCCCTGTCAATTCATGGCTGGGAAACATCATCCAGTACGCTCCAACT 8920

Qy       8581 ATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATGGCTCAAGACACG 8640

Db       8921 ATATGGGTTCGTATGGTCCTGATGACACACTTCTTCTCCATTCTCATGGCCCAAGACACC 8980

Qy       8641 CTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTGAGTCCCTTGGAC 8700

Db       8981 CTGGACCAGAACCTCAACTTTGAAATGTACGGATCGGTGTACTCCGTGAGTCCTTCGGAC 9040

Qy       8701 CTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTGCACACATACACT 8760

Db       9041 CTCCCAGCCATAATTGAAAGGTTACACGGGCTTGACGCCTTCTCTCTGCACACATACACT 9100

Qy       8761 CCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGA 8820

Db       9101 CCCCACGAACTGACGCGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGA 9160

Qy       8821 GCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGTGGGGGGAGAGCG 8880

Db       9161 GCGTGGAAGAGTCGGGCGCGTGCAGTCAGGGCGTCCCTCATCTCCCATGGAGGGAGGGCG 9220

Qy       8881 GCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 8940

Db       9221 GCCGTTTGCGGCCGGTATCTCTTCAACTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 9280

Qy       8941 TTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGCGCCGGCGGGGGC 9000

Db       9281 TTGCCGGAGGCACGCCTCCTGGATTTATCCAGCTGGTTTACCGTCGGCGCCGGCGGGGGC 9340

Qy       9001 GACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTTGGCCTACTCCTA 9060

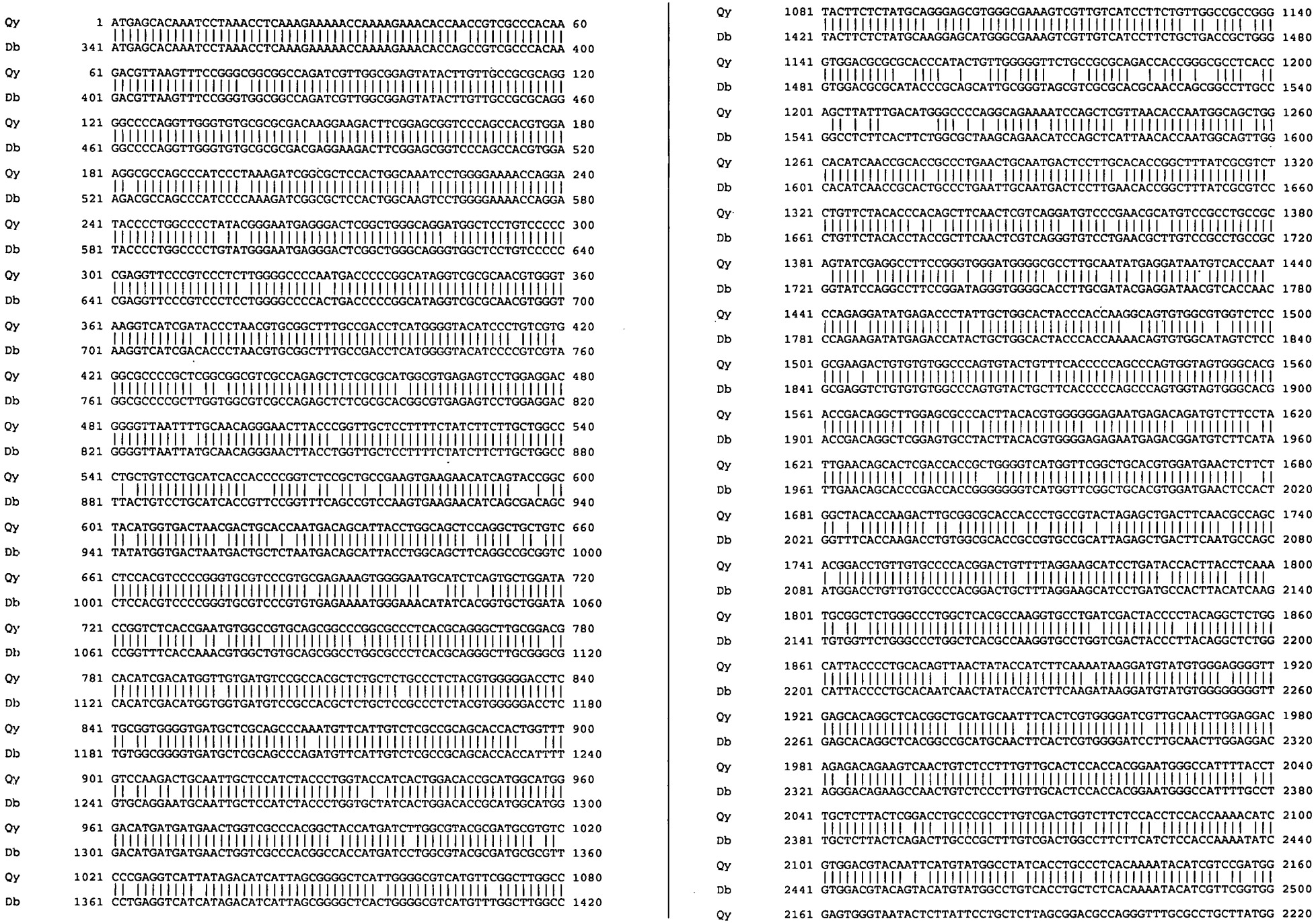
Db       9341 GACATTTATCTCTGCGTGTTGCGTGCCCGACCCCGCTTATTACTCCTTAGCCTACTCCTA 9400

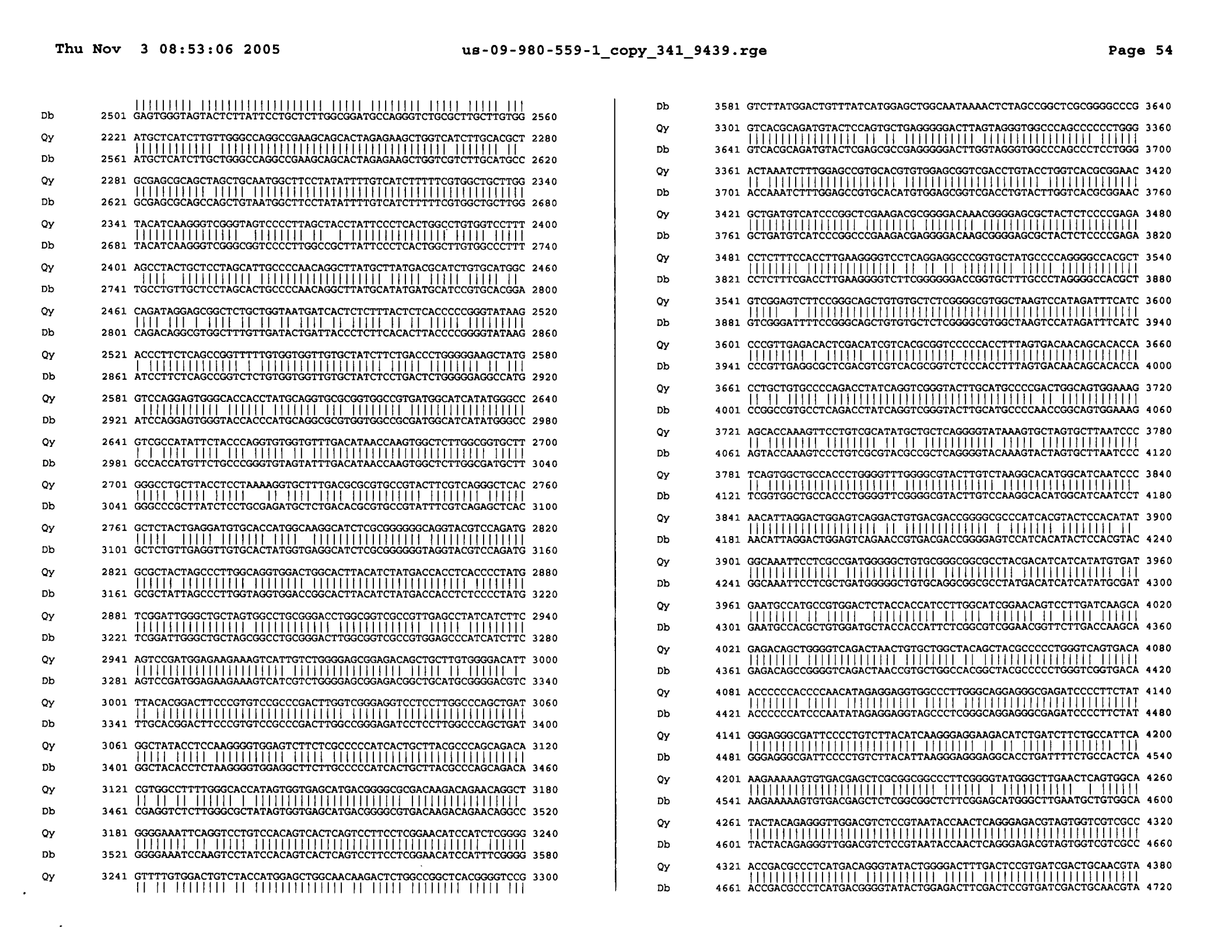
Qy       9061 CTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCG 9098

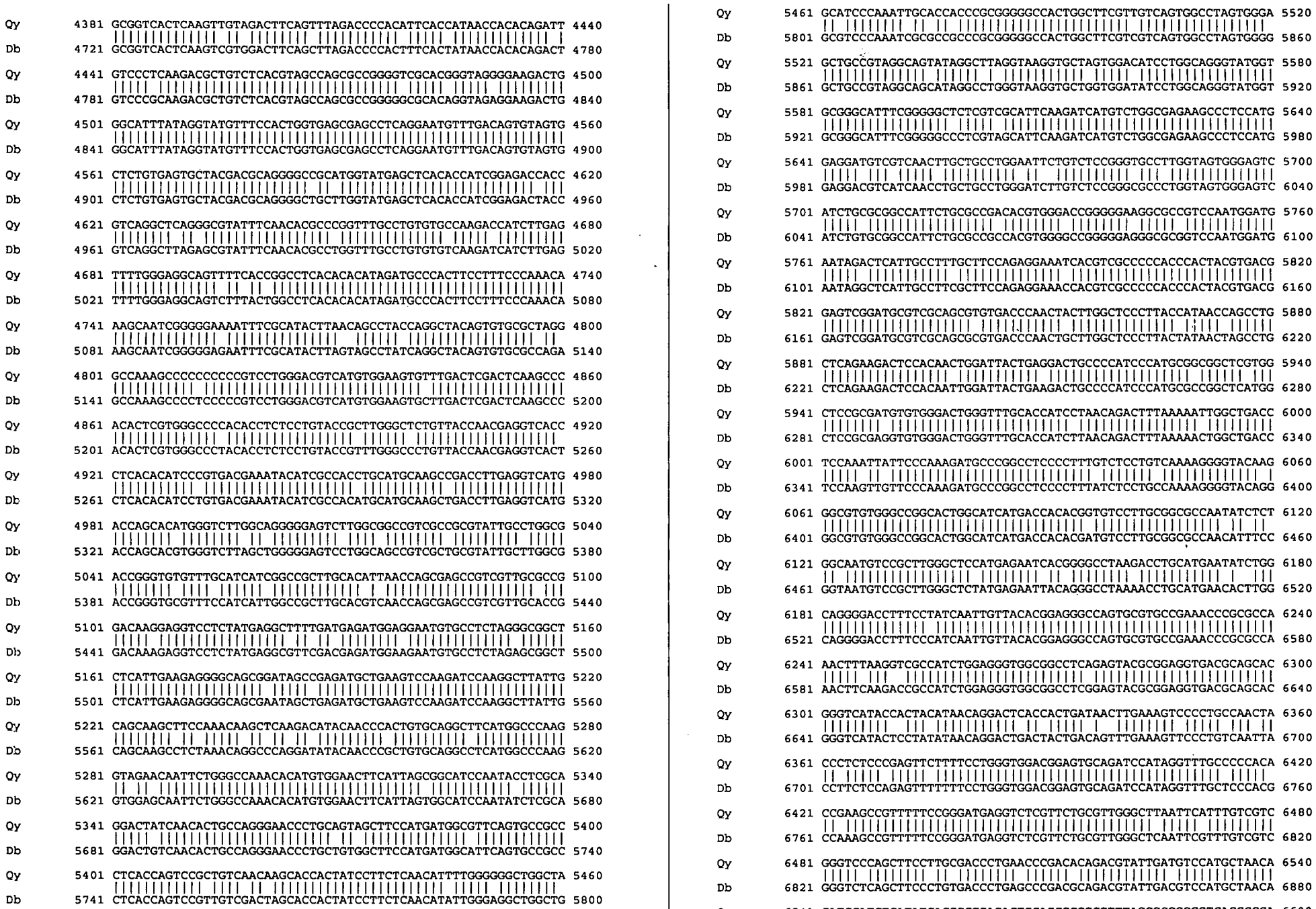
Db       9401 CTTTCCGTGGGACTAGGCCTTTTCCTACTCCCCGCTCG 9438

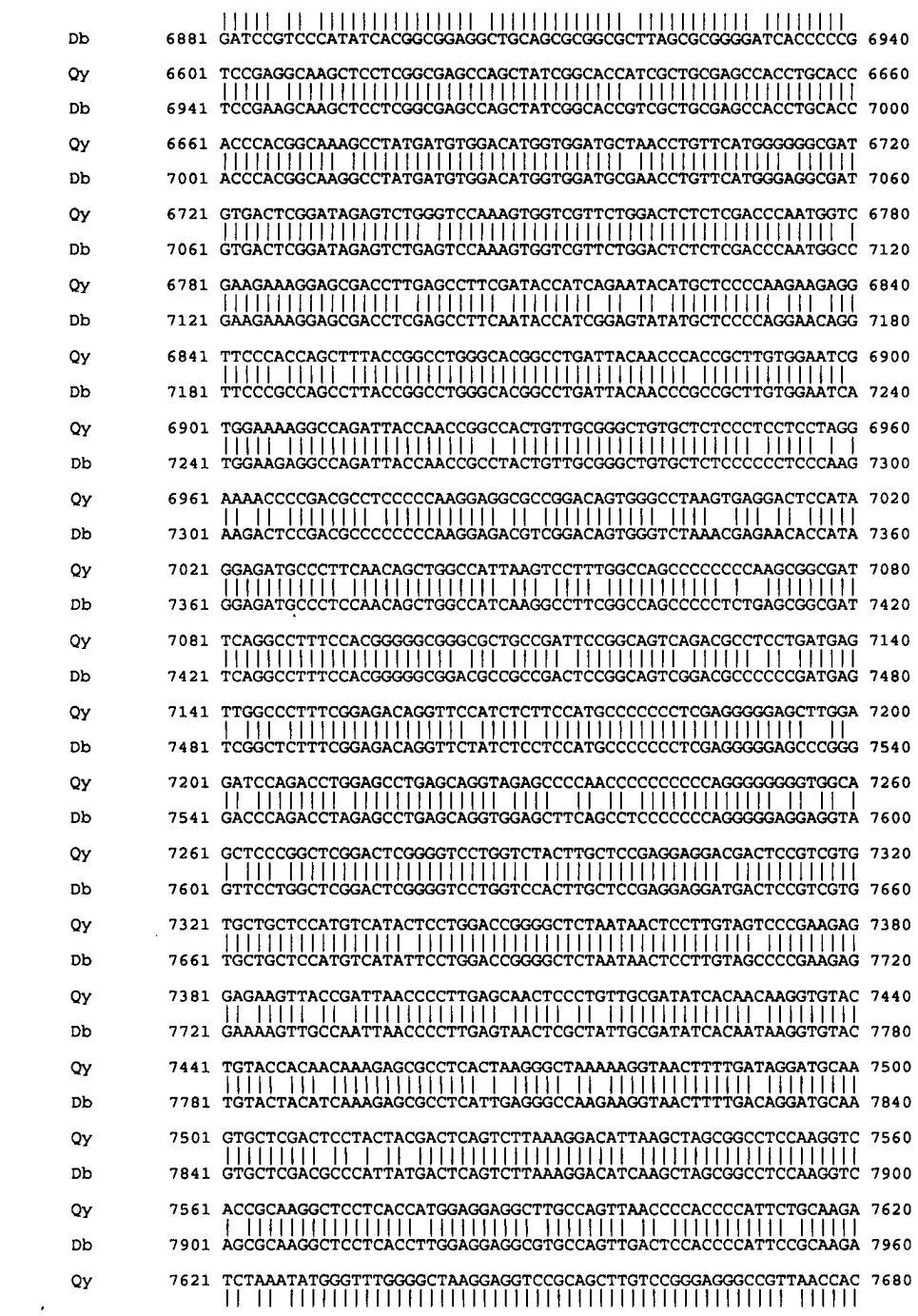

RESULT 12
AF169005
LOCUS       AF169005               9700 bp    RNA     linear   VRL 06-SEP-2001
DEFINITION  Hepatitis C virus isolate NDM59 polyprotein gene, complete cds.
ACCESSION   AF169005
VERSION     AF169005.1  GI:6707285
KEYWORDS    .
SOURCE      Hepatitis C virus
  ORGANISM  Hepatitis C virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepacivirus.
REFERENCE   1  (bases 1 to 9700)
  AUTHORS   Kurihara,C., Ishiyama,N., Nishiyama,Y., Fukushi,S., Kageyama,T.,
            Katayama,K. and Miura,S.
  TITLE     Molecular characterization of hepatitis C virus genotype 2a from
            the entire sequences of four isolates
  JOURNAL   J. Med. Virol. 64 (4), 466-475 (2001)
  MEDLINE   21361470
   PUBMED   11468731
REFERENCE   2  (bases 1 to 9700)
  AUTHORS   Katayama,K. and Kurihara,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-1999) Basic Research Division, BioMedical
            Laboratories, Inc., 1361-1 Matoba, Kawagoe, Saitama 350-1101, Japan
FEATURES             Location/Qualifiers
     source          1. .9700
                     /organism="Hepatitis C virus"
                     /mol_type="genomic RNA"
                     /isolate="NDM59"
                     /db_xref="taxon:11103"
                     /note="genotype: 2a"
     CDS             341. .9442
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="AAF25613.1"
                     /db_xref="GI:6707286"
                     /translation="MSTNPKPQRKTKRNTSRRPQDVKFPGGGQIVGGVYLLPRRGPRL
                     GVRATRKTSERSQPRGRRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRG
                     SRPSWGPTDPRHRSRNVGKVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLED
                     GVNYATGNLPGCSFSIFLLALLSCITVPVSAVQVKNISDSYMVTNDCSNDSITWQLQA
                     AVLHVPGCVPCEKMGNISRCWIPVSPNVAVQRPGALTQGLRAHIDMVVMSATLCSALY
                     VGDLCGGVMLAAQMFIVSPQHHHFVQECNCSIYPGAITGHRMAWDMMMNWSPTATMIL
                     AYAMRVPEVIIDIISGAHWGVMFGLAYFSMQGAWAKVVVILLLTAGVDAHTRSIAGSV
                     AHATSGLAGLFTSGAKQNIQLINTNGSWHINRTALNCNDSLNTGFIASLFYTYRFNSS
                     GCPERLSACRGIQAFRIGWGTLRYEDNVTNPEDMRPYCWHYPPKQCGIVSARSVCGPV
                     YCFTPSPVVVGTTDRLGVPTYTWGENETDVFILNSTRPPGGSWFGCTWMNSTGFTKTC
                     GAPPCRIRADFNASMDLLCPTDCFRKHPDATYIKCGSGPWLTPRCLVDYPYRLWHYPC
                     TINYTIFKIRMYVGGVEHRLTAACNFTRGDPCNLEDRDRSQLSPLLHSTTEWAILPCS
                     YSDLPALSTGLLHLHQNIVDVQYMYGLSPALTKYIVRWEWVVLLFLLLADARVCACLW
                     MLILLGQAEAALEKLVVLHAASAASCNGFLYFVIFFVAAWYIKGRAVPLAAYSLTGLW
                     PFCLLLLALPQQAYAYDASVHGQTGVALLILITLFTLTPGYKILLSRSLWWLCYLLTL
                     GEAMIQEWVPPMQARGGRDGIIWAATMFCPGVVFDITKWLLAMLGPAYLLRDALTRVP
                     YFVRAHALLRLCTMVRHLAGGRYVQMALLALGRWTGTYIYDHLSPMSDWAASGLRDLA
                     VAVEPIIFSPMEKKVIVWGAETAACGDVLHGLPVSARLGREILLGPADGYTSKGWRLL
                     APITAYAQQTRGLLGAIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISGVLWTVYHG
                     AGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRNADVIPA
                     RRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGIFRAAVCSRGVAKSIDFIPVEA
                     LDVVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNPSV
                     AATLGFGAYLSKAHGINPNIRTGVRTVTTGESITYSTYGKFLADGGCAGGAYDIIICD
                     ECHAVDATTILGVGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIP
                     FYGRAIPLSYIKGGRHLIFCHSKKKCDELSAALRSMGLNAVAYYRGLDVSVIPTQGDV
                     VVVATDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQTVPQDAVSRSQRRGR
                     TGRGRLGIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGL
                     PVCQDHLEFWEAVFTGLTHIDAHFLSQTKQSGENFAYLVAYQATVCARAKAPPPSWDV
                     MWKCLTRLKPTLVGPTPLLYRLGPVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAG
                     GVLAAVAAYCLATGCVSIIGRLHVNQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQ
                     RIAEMLKSKIQGLLQQASKQAQDIQPAVQASWPKVEQFWAKHMWNFISGIQYLAGLST
                     LPGNPAVASMMAFSAALTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAA
                     VGSIGLGKVLVDILAGYGAGISGALVAFKIMSGEKPSMEDVINLLPGILSPGALVVGV
                     ICAAILRRHVGPGEGAVQWMNRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTIT
                     SLLRRLHNWITEDCPIPCAGSWLREVWDWVCTILTDFKNWLTSKLFPKMPGLPFISCQ
                     KGYRGVWAGTGIMTTRCPCGANISGNVRLGSMRITGPKTCMNTWQGTFPINCYTEGQC
                     VPKPAPNFKTAIWRVAASEYAEVTQHGSYSYITGLTTDSLKVPCQLPSPEFFSWVDGV
                     QIHRFAPTPKPFFRDEVSFCVGLNSFVVGSQLPCDPEPDADVLTSMLTDPSHITAEAA
                     ARRLARGSPPSEASSSASQLSAPSLRATCTTHGKAYDVDMVDANLFMGGDVTRIESES
                     KVVVLDSLDPMAEERSDLEPSIPSEYMLPRNRFPPALPAWARPDYNPPLVESWKRPDY
                     QPPTVAGCALPPPKKTPTPPPRRRRTVGLNENTIGDALQQLAIKAFGQPPLSGDSGLS
                     TGADAADSGSRTPPDESALSETGSISSMPPLEGEPGDPDLEPEQVELQPPPQGEEVVP
                     GSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEEEKLPINPLSNSLLRYHNKVY
                     CTTSKSASLRAKKVTFDRMQVLDAHYDSVLKDIKLAASKVSARLLTLEEACQLTPPHS
                     ARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSQTPIPTTIMAKNEVFCVDPAKGGK
                     KAARLIVYPDLGVRVCEKMALYDVTQKLPQAVMGASYGFQYSPAQRVEFLLKAWAEKK
                     DPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTERLYVGGPMLNSK
                     GQTCGYRRCRASGVLTTSMGNTITCYIKALAACKAAGIVAPTMLVCGDDLVVISESQG
                     TEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVAXGPQGRRRYYLTR
                     DPTTPLSRAAWETVRHSPVNSWLGNIIQYAPTIWVRMVLMTHFFSILMAQDTLDQNLN
                     FEMYGXVYSVSPLDLPAIIERLHGLDAFSLHTYSPHELTRVASALRKLGAPPLRAWKS
                     RARAVRASLISRGGRAAICGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGGDI
                     YHSVSRARPRLLLLSLLLLVGVGLFLLPAR"
ORIGIN


  Query Match           86.6%;  Score 7878.6;  DB 14;  Length 9700;
  Best Local Similarity  91.6%;  Pred. No. 0;
  Matches 8334;  Conservative    3;  Mismatches  762;  Indels    0;  Gaps    0;
```


Qy 1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 60
Db 341 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAGCCGTCGCCCACAA 400
Qy 61 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 120
Db 401 GACGTTAAGTTTCCGGGTGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 460
Qy 121 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 180
Db 461 GGCCCCAGGTTGGGTGTGCGCGCGACGAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 520
Qy 181 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 240
Db 521 AGACGCCAGCCCATCCCCAAAGATCGGCGCTCCACTGGCAAGTCCTGGGGAAAACCAGGA 580
Qy 241 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 300
Db 581 TACCCCTGGCCCCTGTATGGGAATGAGGGACTCGGCTGGGCAGGGTGGCTCCTGTCCCCC 640
Qy 301 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 360
Db 641 CGAGGTTCCCGTCCCTCCTGGGGCCCCACTGACCCCCGGCATAGGTCGCGCAACGTGGGT 700
Qy 361 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 420
Db 701 AAGGTCATCGACACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCCGTCGTA 760
Qy 421 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 480
Db 761 GGCGCCCCGCTTGGTGGCGTCGCCAGAGCTCTCGCGCACGGCGTGAGAGTCCTGGAGGAC 820
Qy 481 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 540
Db 821 GGGGTTAATTATGCAACAGGGAACTTACCTGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 880
Qy 541 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 600
Db 881 TTACTGTCCTGCATCACCGTTCCGGTTTCAGCCGTCCAAGTGAAGAACATCAGCGACAGC 940
Qy 601 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 660
Db 941 TATATGGTGACTAATGACTGCTCTAATGACAGCATTACCTGGCAGCTTCAGGCCGCGGTC 1000
Qy 661 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 720
Db 1001 CTCCACGTCCCCGGGTGCGTCCCGTGTGAGAAAATGGGAAACATATCACGGTGCTGGATA 1060
Qy 721 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 780
Db 1061 CCGGTTTCACCAAACGTGGCTGTGCAGCGGCCTGGCGCCCTCACGCAGGGCTTGCGGGCG 1120
Qy 781 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 840
Db 1121 CACATCGACATGGTGGTGATGTCCGCCACGCTCTGCTCCGCCCTCTACGTGGGGGACCTC 1180
Qy 841 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 900
Db 1181 TGTGGCGGGGTGATGCTCGCAGCCCAGATGTTCATTGTCTCGCCGCAGCACCACCATTTT 1240
Qy 901 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 960
Db 1241 GTGCAGGAATGCAATTGCTCCATCTACCCTGGTGCTATCACTGGACACCGCATGGCATGG 1300
Qy 961 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1020
Db 1301 GACATGATGATGAACTGGTCGCCCACGGCCACCATGATCCTGGCGTACGCGATGCGCGTT 1360
Qy 1021 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1080
Db 1361 CCTGAGGTCATCATAGACATCATTAGCGGGGCTCACTGGGGCGTCATGTTTGGCTTGGCC 1420
Qy 1081 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1140
Db 1421 TACTTCTCTATGCAAGGAGCATGGGCGAAAGTCGTTGTCATCCTTCTGCTGACCGCTGGG 1480
Qy 1141 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1200
Db 1481 GTGGACGCGCATACCCGCAGCATTGCGGGTAGCGTCGCGCACGCAACCAGCGGCCTTGCC 1540
Qy 1201 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1260
Db 1541 GGCCTCTTCACTTCTGGCGCTAAGCAGAACATCCAGCTCATTAACACCAATGGCAGTTGG 1600
Qy 1261 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1320
Db 1601 CACATCAACCGCACTGCCCTGAATTGCAATGACTCCTTGAACACCGGCTTTATCGCGTCC 1660
Qy 1321 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1380
Db 1661 CTGTTCTACACCTACCGCTTCAACTCGTCAGGGTGTCCTGAACGCTTGTCCGCCTGCCGC 1720
Qy 1381 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1440
Db 1721 GGTATCCAGGCCTTCCGGATAGGGTGGGGCACCTTGCGATACGAGGATAACGTCACCAAC 1780
Qy 1441 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1500
Db 1781 CCAGAAGATATGAGACCATACTGCTGGCACTACCCACCAAAACAGTGTGGCATAGTCTCC 1840
Qy 1501 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1560
Db 1841 GCGAGGTCTGTGTGTGGCCCAGTGTACTGCTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1900
Qy 1561 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1620
Db 1901 ACCGACAGGCTCGGAGTGCCTACTTACACGTGGGGAGAGAATGAGACGGATGTCTTCATA 1960
Qy 1621 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 1680
Db 1961 TTGAACAGCACCCGACCACCGGGGGGGTCATGGTTCGGCTGCACGTGGATGAACTCCACT 2020
Qy 1681 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 1740
Db 2021 GGTTTCACCAAGACCTGTGGCGCACCGCCGTGCCGCATTAGAGCTGACTTCAATGCCAGC 2080
Qy 1741 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 1800
Db 2081 ATGGACCTGTTGTGCCCCACGGACTGCTTTAGGAAGCATCCTGATGCCACTTACATCAAG 2140
Qy 1801 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 1860
Db 2141 TGTGGTTCTGGGCCCTGGCTCACGCCAAGGTGCCTGGTCGACTACCCTTACAGGCTCTGG 2200
Qy 1861 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 1920
Db 2201 CATTACCCCTGCACAATCAACTATACCATCTTCAAGATAAGGATGTATGTGGGGGGGTT 2260
Qy 1921 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 1980
Db 2261 GAGCACAGGCTCACGGCCGCATGCAACTTCACTCGTGGGGATCCTTGCAACTTGGAGGAC 2320
Qy 1981 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2040
Db 2321 AGGGACAGAAGCCAACTGTCTCCCTTGTTGCACTCCACCACGGAATGGGCCATTTTGCCT 2380
Qy 2041 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2100
Db 2381 TGCTCTTACTCAGACTTGCCCGCTTTGTCGACTGGCCTTCTTCATCTCCACCAAAATATC 2440
Qy 2101 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2160
Db 2441 GTGGACGTACAGTACATGTATGGCCTGTCACCTGCTCTCACAAAATACATCGTTCGGTGG 2500
Qy 2161 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2220






Db 6881 GATCCGTCCCATATCACGGCGGAGGCTGCAGCGCGGCGCTTAGCGCGGGGATCACCCCCG 6940
Qy 6601 TCCGAGGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 6660
Db 6941 TCCGAAGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCGTCGCTGCGAGCCACCTGCACC 7000
Qy 6661 ACCCACGGCAAAGCCTATGATGTGGACATGGTGGATGCTAACCTGTTCATGGGGGGCGAT 6720
Db 7001 ACCCACGGCAAGGCCTATGATGTGGACATGGTGGATGCGAACCTGTTCATGGGAGGCGAT 7060
Qy 6721 GTGACTCGGATAGAGTCTGGGTCCAAAGTGGTCGTTCTGGACTCTCTCGACCCAATGGTC 6780
Db 7061 GTGACTCGGATAGAGTCTGAGTCCAAAGTGGTCGTTCTGGACTCTCTCGACCCAATGGCC 7120
Qy 6781 GAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCAGAATACATGCTCCCCAAGAAGAGG 6840
Db 7121 GAAGAAAGGAGCGACCTCGAGCCTTCAATACCATCGGAGTATATGCTCCCCAGGAACAGG 7180
Qy 6841 TTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCGCTTGTGGAATCG 6900
Db 7181 TTCCCGCCAGCCTTACCGGCCTGGGCACGGCCTGATTACAACCCGCCGCTTGTGGAATCA 7240
Qy 6901 TGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTCCCTCCTCCTAGG 6960
Db 7241 TGGAAGAGGCCAGATTACCAACCGCCTACTGTTGCGGGCTGTGCTCTCCCCCCTCCCAAG 7300
Qy 6961 AAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGTGAGGACTCCATA 7020
Db 7301 AAGACTCCGACGCCCCCCCCAAGGAGACGTCGGACAGTGGGTCTAAACGAGAACACCATA 7360
Qy 7021 GGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCCCCAAGCGGCGAT 7080
Db 7361 GGAGATGCCCTCCAACAGCTGGCCATCAAGGCCTTCGGCCAGCCCCCTCTGAGCGGCGAT 7420
Qy 7081 TCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACGCCTCCTGATGAG 7140
Db 7421 TCAGGCCTTTCCACGGGGGCGGACGCCGCCGACTCCGGCAGTCGGACGCCCCCCGATGAG 7480
Qy 7141 TTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCTCGAGGGGGAGCTTGGA 7200
Db 7481 TCGGCTCTTTCGGAGACAGGTTCTATCTCCTCCATGCCCCCCCTCGAGGGGGAGCCCGGG 7540
Qy 7201 GATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCCAGGGGGGGGTGGCA 7260
Db 7541 GACCCAGACCTAGAGCCTGAGCAGGTGGAGCTTCAGCCTCCCCCCCAGGGGGAGGAGGTA 7600
Qy 7261 GCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGACGACTCCGTCGTG 7320
Db 7601 GTTCCTGGCTCGGACTCGGGGTCCTGGTCCACTTGCTCCGAGGAGGATGACTCCGTCGTG 7660
Qy 7321 TGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGTAGTCCCGAAGAG 7380
Db 7661 TGCTGCTCCATGTCATATTCCTGGACCGGGGCTCTAATAACTCCTTGTAGCCCCGAAGAG 7720
Qy 7381 GAGAAGTTACCGATTAACCCCTTGAGCAACTCCCTGTTGCGATATCACAACAAGGTGTAC 7440
Db 7721 GAAAAGTTGCCAATTAACCCCTTGAGTAACTCGCTATTGCGATATCACAATAAGGTGTAC 7780
Qy 7441 TGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTTGATAGGATGCAA 7500
Db 7781 TGTACTACATCAAAGAGCGCCTCATTGAGGGCCAAGAAGGTAACTTTTGACAGGATGCAA 7840
Qy 7501 GTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCGGCCTCCAAGGTC 7560
Db 7841 GTGCTCGACGCCCATTATGACTCAGTCTTAAAGGACATCAAGCTAGCGGCCTCCAAGGTC 7900
Qy 7561 ACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCCCATTCTGCAAGA 7620
Db 7901 AGCGCAAGGCTCCTCACCTTGGAGGAGGCGTGCCAGTTGACTCCACCCCATTCCGCAAGA 7960
Qy 7621 TCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGGGCCGTTAACCAC 7680
Db 7961 TCCAAGTATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGGGCCGTCAACCAC 8020
Qy 7681 ATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATTCCCACAACCATT 7740
Db 8021 ATCAAGTCCGTGTGGAAGGACCTCCTGGAAGACTCACAAACACCAATTCCTACGACCATC 8080
Qy 7741 ATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAGAAAGCAGCTCGC 7800
Db 8081 ATGGCCAAAAATGAGGTGTTCTGCGTGGACCCTGCCAAGGGGGGCAAGAAAGCAGCTCGC 8140
Qy 7801 CTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCCCTTTATGACATT 7860
Db 8141 CTTATCGTCTACCCTGACCTCGGCGTCAGGGTTTGCGAGAAGATGGCCCTTTATGATGTC 8200
Qy 7861 ACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAGTATTCCCCCGCT 7920
Db 8201 ACACAAAAGCTTCCTCAGGCGGTGATGGGGGCTTCCTACGGCTTCCAGTACTCCCCCGCT 8260
Qy 7921 CAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCTATGGGTTTTTCG 7980
Db 8261 CAACGGGTGGAGTTTCTCTTAAAGGCATGGGCGGAAAAGAAAGACCCTATGGGTTTTTCG 8320
Qy 7981 TATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGGACTGAGGAGTCC 8040
Db 8321 TATGATACCCGATGCTTTGATTCAACCGTCACTGAGAGAGACATTAGAACTGAGGAGTCC 8380
Qy 8041 ATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACACTCGCTAACTGAG 8100
Db 8381 ATATACCGGGCCTGCTCCCTGCCCGAGGAGGCCCACACTGCCATACACTCGCTAACTGAG 8440
Qy 8101 AGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGCGGGTACAGGCGT 8160
Db 8441 AGACTCTACGTGGGAGGGCCCATGCTCAACAGCAAGGGTCAGACCTGCGGATACAGGCGT 8500
Qy 8161 TGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACATGCTACGTGAAA 8220
Db 8501 TGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACATGCTACATAAAA 8560
Qy 8221 GCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTGGTATGCGGCGAT 8280
Db 8561 GCCCTAGCGGCTTGCAAGGCTGCGGGGATAGTTGCGCCTACCATGCTGGTATGCGGCGAC 8620
Qy 8281 GACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGGAACCTGAGAGCC 8340
Db 8621 GACTTGGTTGTTATCTCAGAAAGCCAGGGGACTGAGGAGGACGAGCGAAACCTGAGAGCC 8680
Qy 8341 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCCCAGACCGGAGTAT 8400
Db 8681 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCCCAGACCGGAATAT 8740
Qy 8401 GATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTGGGCCCACAAGGC 8460
Db 8741 GACCTGGAGCTGATAACATCTTGTTCCTCAAATGTATCCGTGGCARTCGGCCCACAGGGC 8800
Qy 8461 CGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGGGCTGCCTGGGAA 8520
Db 8801 CGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCACTCTCCCGGGCTGCCTGGGAA 8860
Qy 8521 ACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAGTACGCCCCGACC 8580
Db 8861 ACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAGTACGCCCCAACC 8920
Qy 8581 ATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATGGCTCAAGACACG 8640
Db 8921 ATATGGGTKCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATGGCTCAAGACACC 8980
Qy 8641 CTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTGAGTCCCTTGGAC 8700
Db 8981 CTAGACCAGAACCTCAACTTTGAGATGTACGGGKCGGTCTACTCCGTGAGTCCTTTGGAC 9040
Qy 8701 CTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTGCACACATACACT 8760
Db 9041 CTCCCAGCCATAATTGAAAGGTTACACGGGCTTGACGCTTTTTCTCTGCACACATACTCT 9100

```
Qy       8761 CCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGA 8820
             |||||||| ||||| ||||||||||:||||||||||||||||||||||||||||||||||
Db       9101 CCCCACGAGCTGACGCGGGTGGCTTCRGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGA 9160

Qy       8821 GCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGTGGGGGGAGAGCG 8880
             |||||||||||| ||||| |||||||| |||||||||||||||||||||| |||||||||
Db       9161 GCGTGGAAGAGTCGGGCGCGTGCAGTTAGGGCGTCCCTCATCTCCCGTGGAGGGAGAGCG 9220

Qy       8881 GCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 8940
             ||| |||||||||| |||||||||| ||||||||||||||||||||||||||||||||||
Db       9221 GCCATTTGCGGTCGGTATCTCTTCAACTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 9280

Qy       8941 TTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGCGCCGGCGGGGGC 9000
             |||||||| || |||||||||||||||||||| ||||| ||||||||||||||||||||||
Db       9281 TTGCCGGAGGCACGCCTCCTGGATTTATCCAGTTGGTTTACCGTCGGCGCCGGCGGGGGC 9340

Qy       9001 GACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTTGGCCTACTCCTA 9060
             |||||||||||||||||||||| |||||||||||||||| ||| || ||||||||||||
Db       9341 GACATTTATCACAGCGTGTCGCGCGCCCGACCCCGCTTATTACTCCTTAGCCTACTCCTA 9400

Qy       9061 CTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCGG 9099
             ||| ||||||||||||||||||||||||||||||||||||
Db       9401 CTTCTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCGG 9439
```

```
RESULT 13
AF238482
LOCUS       AF238482               9416 bp    RNA     linear   VRL 26-MAR-2000
DEFINITION  Hepatitis C virus 2a polyprotein gene, complete cds.
ACCESSION   AF238482
VERSION     AF238482.1  GI:7329202
KEYWORDS    .
SOURCE      Hepatitis C virus
  ORGANISM  Hepatitis C virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepacivirus.
REFERENCE   1  (bases 1 to 9416)
  AUTHORS   Itakura,J., Nagayama,K., Enomoto,N., Kurosaki,M., Watanabe,H. and
            Sato,C.
  TITLE     Full length cDNA sequence of HCV genotype 2a, strain MD2a-2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 9416)
  AUTHORS   Itakura,J., Nagayama,K., Enomoto,N., Kurosaki,M., Watanabe,H. and
            Sato,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-FEB-2000) Second Department of Internal Medicine,
            Tokyo Medical and Dental University, 1-5-45 Yushima Bunkyo-Ku,
            Tokyo 113-8519, Japan
FEATURES             Location/Qualifiers
     source          1. .9416
                     /organism="Hepatitis C virus"
                     /mol_type="genomic RNA"
                     /strain="MD2a-2"
                     /db_xref="taxon:11103"
                     /note="genotype: 2a"
     CDS             307. .9408
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="AAF59941.1"
                     /db_xref="GI:7329203"
                     /translation="MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRL
                     GVRATRKTSERSQPRGRRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRG
                     SRPSWGPTDPRHRSRNVGKVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLED
                     GVNYATGNLPGCSFSIFLLALLSCVTVPVSAVQVKNTSETYMVTNDCSNDSITWQLQA
                     AVLHVPGCVPCERTGNTSRCWIPVSPNVAVQQPGALTQGLRTHIDMVVMSATLCSALY
                     VGDLCGGVVLAAQLFIVSPRRHWFVQECNCSIYPGAITGHRMAWDMMMNWSPTTTMVL
                     AYAMRVPEVIIDIISGAHWGVMFGLAYFSMQGAWAKVAVILLLTAGVEARTHTTGSVA
                     GRTTSGFAGIFTSGPKQNIQLINTNGSWHINRTALNCNDSLNTGFMAALFYTKNFNSS
                     GCPERLSXCRNIEAFRIGWGTLQYEDDVTNPEDMRPYCWHYPPKQCGIFPAGSVCGPV
                     YCFTPSPVVVGTTNKLGVPTYTWGENETDVFILNSTRPPRGSWFGCTWMNSTGFTKTC
                     GAPPCRIRADFNASTDLLCPTDCFRKHPEATYIKCGSGPWLTPRCLVDYPYRLWHYPC
                     TVNYSIFKIRMYVGGVEHRLTAACNFSRGDRCNLEDRDRSQLTPLLHSTTEWAILPCS
                     YSDLPALSTGLLHLHQNIVDVQYMYGLTPALTKYVVRWEWVVLLFLLLADARVCACLW
                     MLILLGQAEAALEKLVVLHAASAASCNGLLYFVIFFVAAWYIKGRAVPLAAYSLTGLW
                     PFCLLLLVLPQQAYAYDASVHGQVGAALLILITLFTLTPGYKTLLSRCLWWLCYLLTL
                     GEALIQEWVPPTQARGGRDGIIWAATMFCPGVVFDITKWLLAVLGPAYLLRDALTRVP
                     YFVRAHALLRMCTVVRNLAGGKYVQMALLALGRWTGTYIYDHLTPMSDWAASGLRDLA
                     VAVEPIIFSPMEKKVIAWGAETAACGDIXHGLPVSARLGREILLGPADGYTSKGWRLL
                     APITAYAQQTRGLLGAIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTTISGVLWTVYHG
                     AGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRNADVIPA
                     RRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGIFRAAVCSRGVAKSIDFIPVET
                     LDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNPSV
                     AATLGFGAYLSKAHGINPNIRTGVRTVTTGAAITYSTYGKFLADGGCAGGAYDIIICD
                     ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIP
                     FYGRAIPLSHIKGGRHLIFCHSKKKCDELAAALRGMGLNAVAYYRGLDVSVIPTQGDV
                     VVFATDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQTVPQDAVSRSQRRGR
                     TGRGRLGIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPAETTVRLRAYFNTPGL
                     PVCQDHLEFWEAVFTGLTHIDAHFLSQTKQSGENFAYLVAYQATVCARAKAPPPSWDV
                     MWKCLTRLKPTLVGPTPLLYRLGSVANEITLTHPVTKYIATCMQADLEVMTSTWVLAG
                     GVLAAVAAYCLATGCVSIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALVEEGQ
                     RIAEMLKSKIQGLLQQASKQAQDIQPAVQASWPKVEQFWAKHMWNFISGIQYLAGLST
                     LPGNPAVASMMAFSAALTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAA
                     VGSIGLGKVLVDILAGYGAGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGV
                     ICAAILRRHVGPGEGAVQWMNRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTIT
                     SLLRRLHNWITEDCPIPCAGSWLRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFISCQ
                     KGYKGVWAGTGIMTTRCPCGANISGNVRLGSMRITGPKTCMNTWQGTFPINCYTEGQC
                     VPKPAPNFKVAIWRVAASEYAEVTQHGSYSYITGLTTDNLKVPCQLPSPEFFSWVDGV
                     QIHRFAPTPKPFFRDEVSFCVGLNSFVVGSQLPCDPEPDTDVLTSMLTDPSHITAETA
                     ARRLARGSPPSEASSSASQLSAPSLRATCTTHGKTYDVDMVDASLFMGGDVTRIESES
                     KVVVLDSLDPMAEEESDLEPSIPSEYMLSRNRFPPALPAWARPDYNPPLVESWKRPDY
                     QPATVAGCALPPPKKAPTPPPRRRRTVGLSESTIGDALQQLAIKSFGQPPPSGDSGLS
                     TGADAADSGSRSPPDELALSETGSTSSMPPLEGEPGDPDLEPEQVEPQPPPQGGEVAP
                     GSDSGSWSTCSEENVSVVFCSISYSWTGGLITPCIPEKEKLPFNPLSNSLLRYDNKVY
                     CTTSKSASLRAKKVTFDRMQVLDAYYDSVLKDIKLAASKVSARLLTLEEACQLTPPHS
                     ARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSQTPIPTTIMAKNEVFCVDPNKGGK
                     KAARLIVYPDLGVRVCEKMALYDVTQKLPQAVMGASYGFQYSPAQRVEFLLKAWAEKK
                     DPMGFSYDTRCFDSTVTERDIRTEESIYRPCSLPEEAHTAIHSLTERLYVGGPMFNSK
                     GQACGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIVAPTMLVCGDDLVVISESQG
                     TEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVHLGPQGRRRYYLTR
                     DPTTPIARAAWETVRHSPVNSWLGNIIQYAPTVWVRMVLMTHFFSILMAQDTLDQNLN
                     FEMYGSVYSVSPLDLPAIIERLHGLDAFSLHTYSPHELTRVAAALRKLGAPPLRAWKS
                     RARAVRASLISHGGRAAVCGRYLFNWAVKTKLKLTPLPEARHLDLSSWFTVGAGGHI
                     YHSVSRARPRLLLLGLLLLFVGVGLFLLPAR"
ORIGIN

  Query Match           86.5%;  Score 7869.8;  DB 14;  Length 9416;
  Best Local Similarity  91.5%;  Pred. No. 0;
  Matches 8330;  Conservative    0;  Mismatches  769;  Indels    0;  Gaps    0;

Qy          1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 60
              |||||||||||||||||||||||||||||||||||||||||||| || ||||||||||||||
Db        307 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAATACTAACCGTCGCCCACAG 366

Qy         61 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 120
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        367 GACGTCAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 426

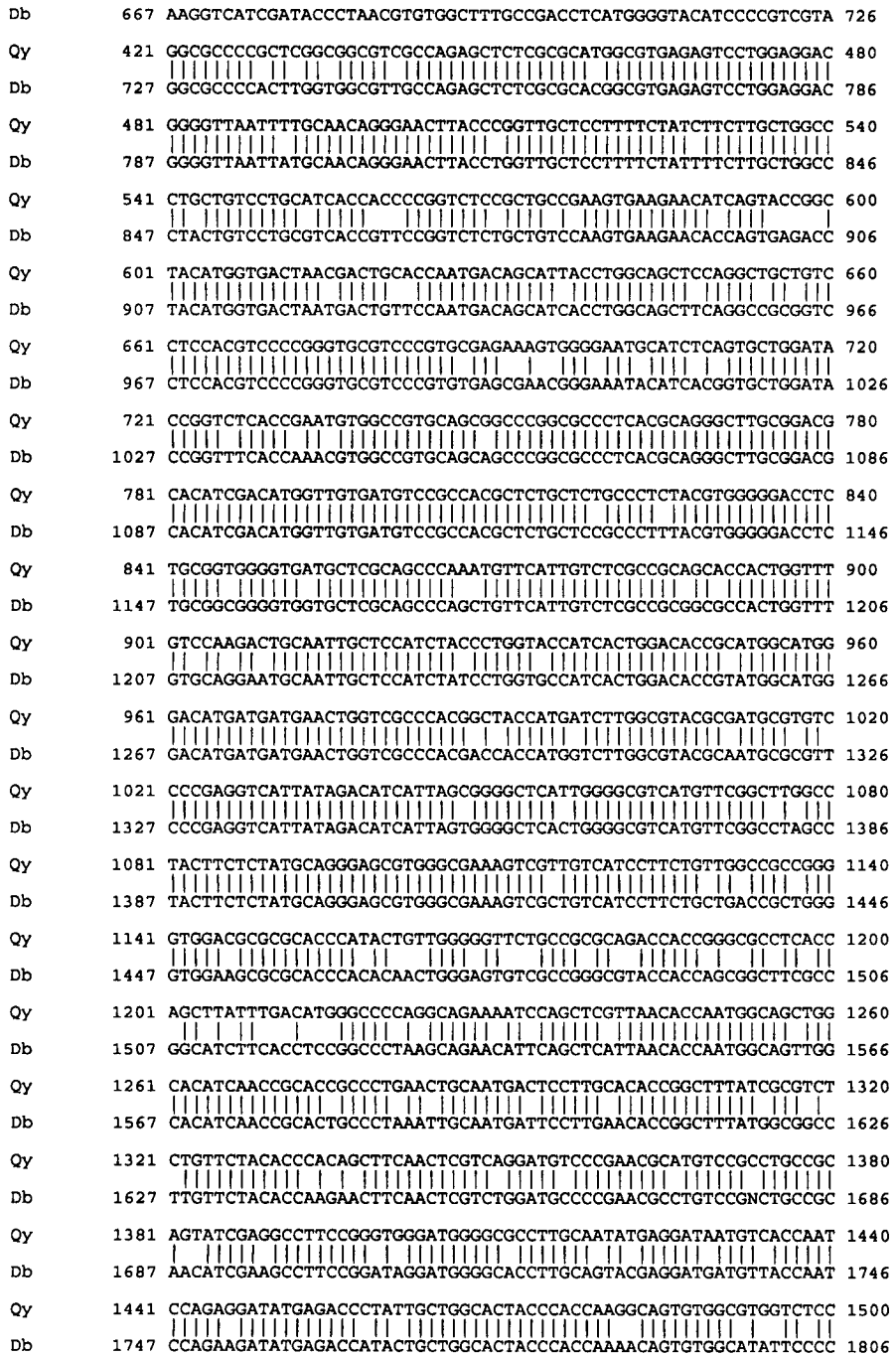
Qy        121 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        427 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGG 486

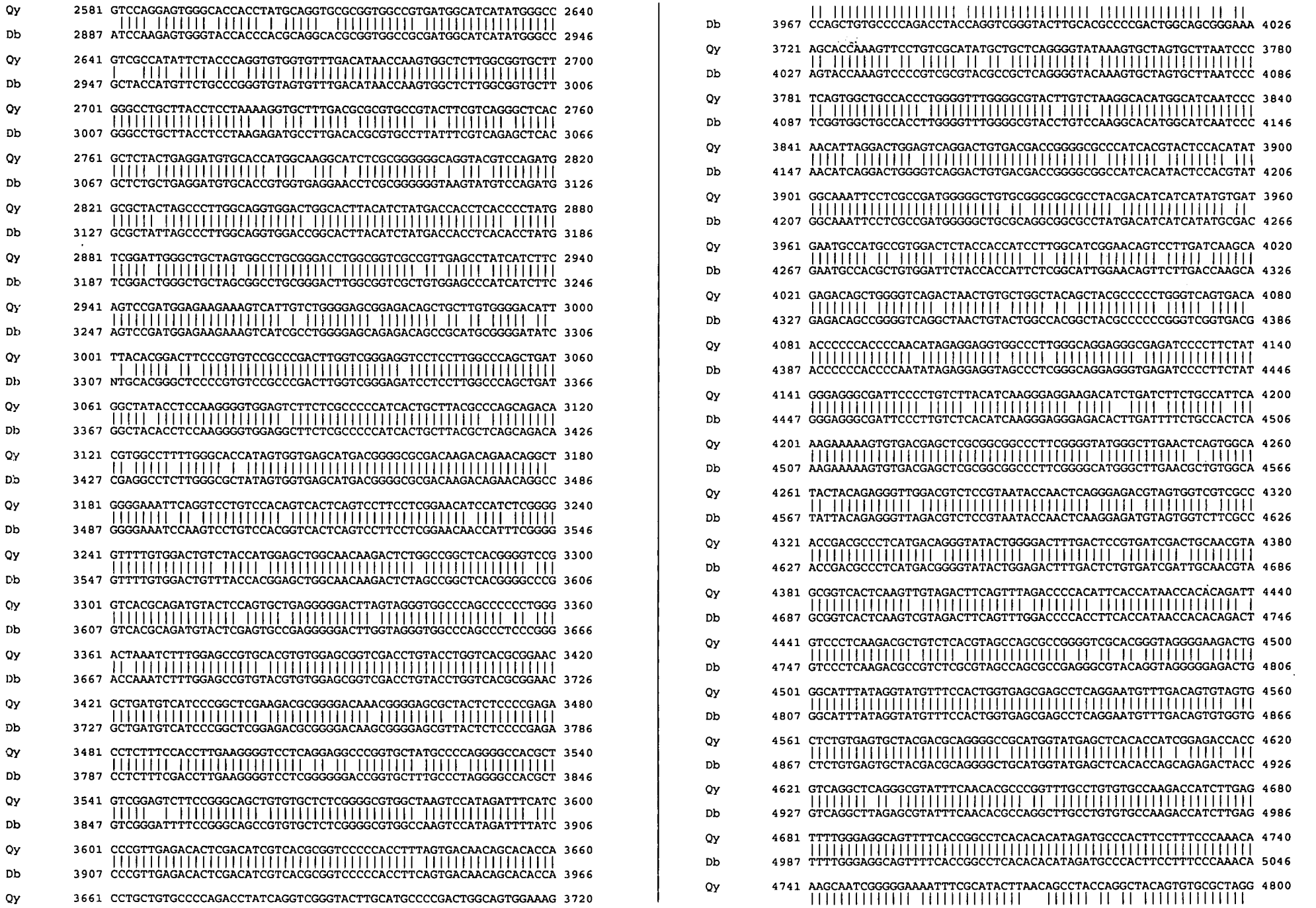
Qy        181 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 240
              |||||||||||||||| |||||||||||||||||||| |||||||||||||||||||||
Db        487 AGGCGCCAGCCCATCCCCAAAGATCGGCGCTCCACTGGCAAGTCCTGGGGAAAACCAGGA 546

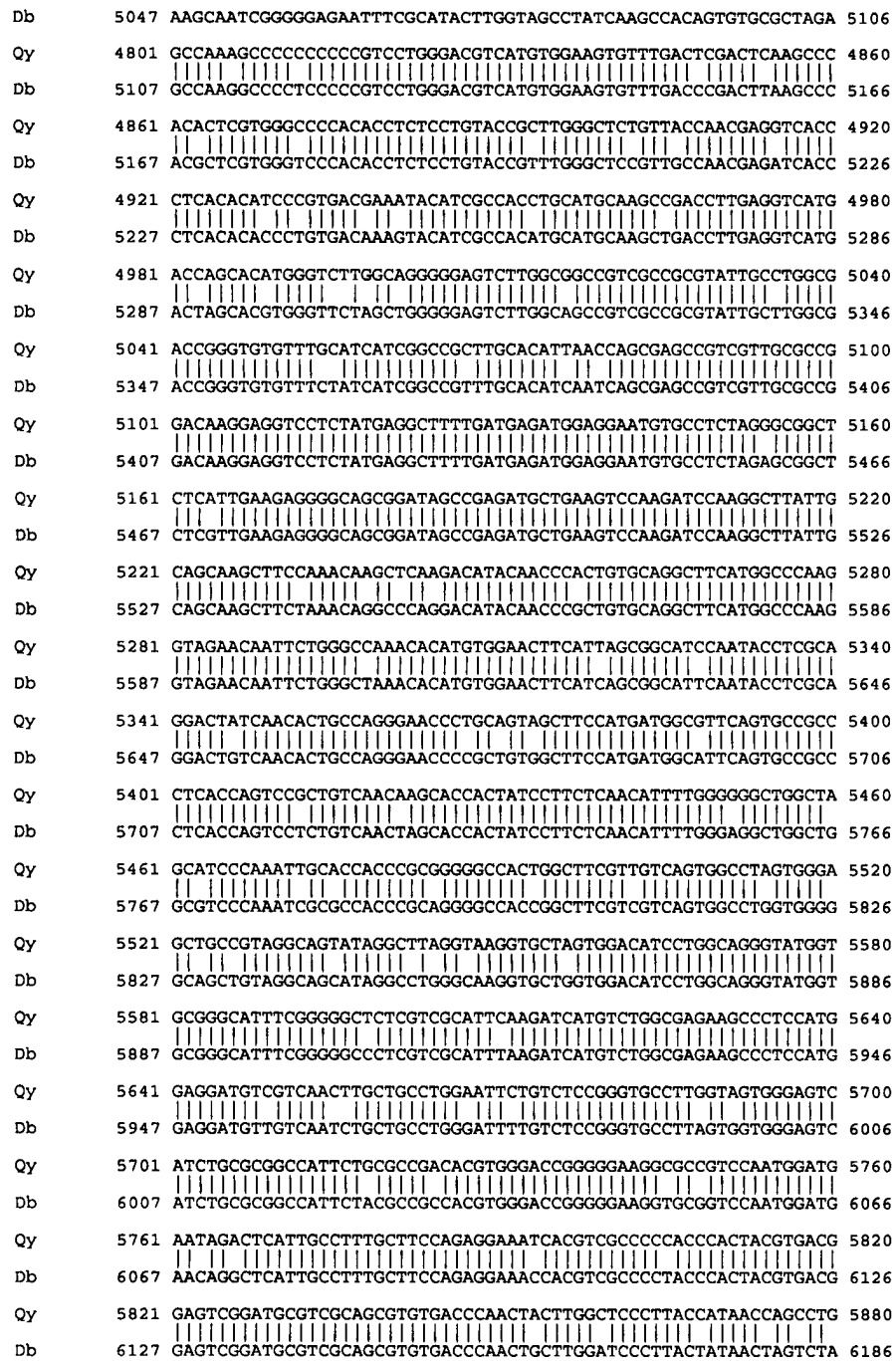
Qy        241 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 300
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        547 TACCCCTGGCCCCTATATGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 606

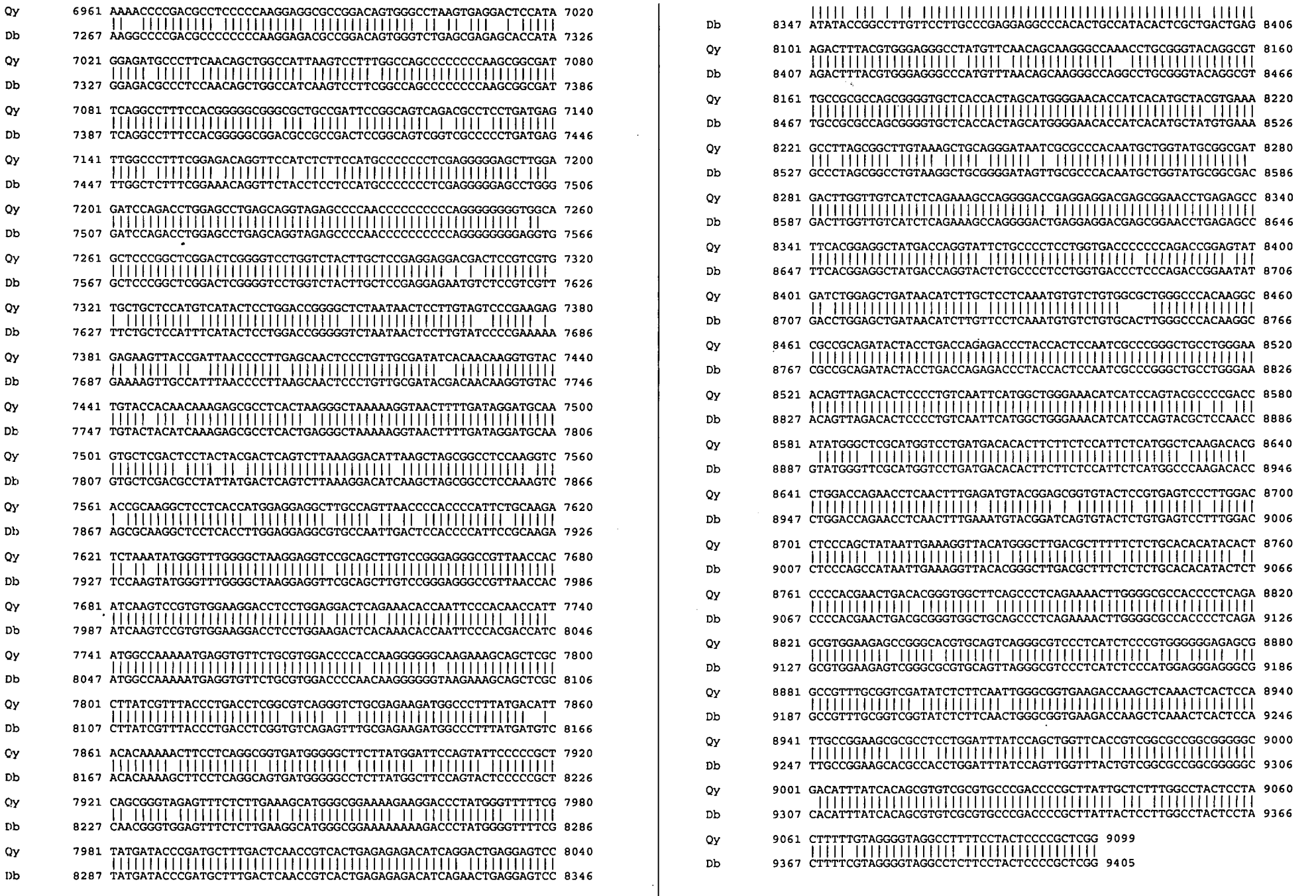
Qy        301 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 360
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        607 CGAGGTTCCCGTCCCTCTTGGGGCCCCACTGACCCCCGGCATAGGTCGCGCAACGTGGGT 666

Qy        361 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 420
              |||||||||||||||||||||||| |||||||||||||||||||||||||||| |||||
```


Db 667 AAGGTCATCGATACCCTAACGTGTGGCTTTGCCGACCTCATGGGGTACATCCCCGTCGTA 726
Qy 421 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 480
Db 727 GGCGCCCCACTTGGTGGCGTTGCCAGAGCTCTCGCGCACGGCGTGAGAGTCCTGGAGGAC 786
Qy 481 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 540
Db 787 GGGGTTAATTATGCAACAGGGAACTTACCTGGTTGCTCCTTTTCTATTTTCTTGCTGGCC 846
Qy 541 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 600
Db 847 CTACTGTCCTGCGTCACCGTTCCGGTCTCTGCTGTCCAAGTGAAGAACACCAGTGAGACC 906
Qy 601 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 660
Db 907 TACATGGTGACTAATGACTGTTCCAATGACAGCATCACCTGGCAGCTTCAGGCCGCGGTC 966
Qy 661 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 720
Db 967 CTCCACGTCCCCGGGTGCGTCCCGTGTGAGCGAACGGGAAATACATCACGGTGCTGGATA 1026
Qy 721 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 780
Db 1027 CCGGTTTCACCAAACGTGGCCGTGCAGCAGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 1086
Qy 781 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 840
Db 1087 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCCGCCCTTTACGTGGGGGACCTC 1146
Qy 841 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 900
Db 1147 TGCGGCGGGGTGGTGCTCGCAGCCCAGCTGTTCATTGTCTCGCCGCGGCGCCACTGGTTT 1206
Qy 901 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 960
Db 1207 GTGCAGGAATGCAATTGCTCCATCTATCCTGGTGCCATCACTGGACACCGTATGGCATGG 1266
Qy 961 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1020
Db 1267 GACATGATGATGAACTGGTCGCCCACGACCACCATGGTCTTGGCGTACGCAATGCGCGTT 1326
Qy 1021 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1080
Db 1327 CCCGAGGTCATTATAGACATCATTAGTGGGGCTCACTGGGGCGTCATGTTCGGCCTAGCC 1386
Qy 1081 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1140
Db 1387 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGCTGTCATCCTTCTGCTGACCGCTGGG 1446
Qy 1141 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1200
Db 1447 GTGGAAGCGCGCACCCACACAACTGGGAGTGTCGCCGGGCGTACCACCAGCGGCTTCGCC 1506
Qy 1201 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1260
Db 1507 GGCATCTTCACCTCCGGCCCTAAGCAGAACATTCAGCTCATTAACACCAATGGCAGTTGG 1566
Qy 1261 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1320
Db 1567 CACATCAACCGCACTGCCCTAAATTGCAATGATTCCTTGAACACCGGCTTTATGGCGGCC 1626
Qy 1321 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1380
Db 1627 TTGTTCTACACCAAGAACTTCAACTCGTCTGGATGCCCCGAACGCCTGTCCGNCTGCCGC 1686
Qy 1381 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1440
Db 1687 AACATCGAAGCCTTCCGGATAGGATGGGGCACCTTGCAGTACGAGGATGATGTTACCAAT 1746
Qy 1441 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1500
Db 1747 CCAGAAGATATGAGACCATACTGCTGGCACTACCCACCAAAACAGTGTGGCATATTCCCC 1806
Qy 1501 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1560
Db 1807 GCGGGGTCTGTGTGCGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTAGTAGTGGGCACA 1866
Qy 1561 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1620
Db 1867 ACCAACAAACTTGGAGTGCCCACTTACACATGGGGGGAGAATGAGACAGATGTCTTCATA 1926
Qy 1621 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 1680
Db 1927 TTGAACAGCACCCGACCACCGCGGGGGTCATGGTTCGGCTGCACGTGGATGAATTCCACT 1986
Qy 1681 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 1740
Db 1987 GGTTTCACCAAGACTTGTGGCGCACCACCTTGCCGCATTAGGGCCGACTTCAATGCCAGC 2046
Qy 1741 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 1800
Db 2047 ACGGACCTCTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGAAGCCACTTACATCAAA 2106
Qy 1801 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 1860
Db 2107 TGTGGTTCTGGGCCCTGGCTCACGCCAAGGTGCCTGGTCGACTACCCCTACAGGCTCTGG 2166
Qy 1861 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 1920
Db 2167 CATTACCCCTGCACAGTCAACTATTCCATCTTTAAGATAAGGATGTATGTGGGGGGGGTT 2226
Qy 1921 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 1980
Db 2227 GAGCACAGGCTCACGGCCGCGTGCAATTTCTCTCGTGGGGATCGTTGCAACTTGGAGGAC 2286
Qy 1981 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2040
Db 2287 AGGGACAGGAGTCAACTGACTCCTCTGTTGCACTCCACCACGGAATGGGCCATCCTTCCC 2346
Qy 2041 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2100
Db 2347 TGTTCTTACTCAGACTTGCCCGCTTTGTCGACTGGCCTTCTCCATCTCCACCAAAATATC 2406
Qy 2101 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2160
Db 2407 GTGGACGTACAATACATGTATGGCCTGACACCTGCCCTCACAAAATATGTCGTTCGATGG 2466
Qy 2161 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2220
Db 2467 GAGTGGGTAGTACTCTTATTCCTGCTCCTAGCGGACGCCAGGGTTTGCGCCTGCTTGTGG 2526
Qy 2221 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2280
Db 2527 ATGCTCATCTTGCTGGGCCAGGCTGAAGCAGCACTAGAGAAGCTGGTCGTCTTGCACGCT 2586
Qy 2281 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATATTTTGTCATCTTTTTCGTGGCTGCTTGG 2340
Db 2587 GCGAGTGCGGCTAGCTGCAATGGCCTCCTGTATTTTGTCATCTTTTTCGTGGCTGCTTGG 2646
Qy 2341 TACATCAAGGGTCGGGTAGTCCCCTTAGCTACCTATTCCCTCACTGGCCTGTGGTCCTTT 2400
Db 2647 TACATCAAGGGTCGGGCGGTCCCCTTGGCCGCCTACTCCCTCACTGGCCTATGGCCCTTC 2706
Qy 2401 AGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTGCATGGC 2460
Db 2707 TGCCTACTACTCCTAGTACTGCCCCAACAGGCTTATGCCTATGATGCATCCGTGCACGGA 2766
Qy 2461 CAGATAGGAGCGGCTCTGCTGGTAATGATCACTCTCTTTACTCTCACCCCCGGGTATAAG 2520
Db 2767 CAGGTAGGCGCGGCTTTGTTAATACTGATTACCCTCTTCACACTTACCCCGGGGTATAAG 2826
Qy 2521 ACCCTTCTCAGCCGGTTTTTGTGGTGGTTGTGCTATCTTCTGACCCTGGGGGAAGCTATG 2580
Db 2827 ACCCTCCTCAGTCGGTGTCTGTGGTGGTTGTGCTATCTCCTGACCCTGGGGGAAGCCTTG 2886




Db 5047 AAGCAATCGGGGGAGAATTTCGCATACTTGGTAGCCTATCAAGCCACAGTGTGCGCTAGA 5106
Qy 4801 GCCAAAGCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 4860
Db 5107 GCCAAGGCCCCTCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACCCGACTTAAGCCC 5166
Qy 4861 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 4920
Db 5167 ACGCTCGTGGGTCCCACACCTCTCCTGTACCGTTTGGGCTCCGTTGCCAACGAGATCACC 5226
Qy 4921 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 4980
Db 5227 CTCACACACCCTGTGACAAAGTACATCGCCACATGCATGCAAGCTGACCTTGAGGTCATG 5286
Qy 4981 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5040
Db 5287 ACTAGCACGTGGGTTCTAGCTGGGGGAGTCTTGGCAGCCGTCGCCGCGTATTGCTTGGCG 5346
Qy 5041 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5100
Db 5347 ACCGGGTGTGTTTCTATCATCGGCCGTTTGCACATCAATCAGCGAGCCGTCGTTGCGCCG 5406
Qy 5101 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5160
Db 5407 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGAGCGGCT 5466
Qy 5161 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5220
Db 5467 CTCGTTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5526
Qy 5221 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5280
Db 5527 CAGCAAGCTTCTAAACAGGCCCAGGACATACAACCCGCTGTGCAGGCTTCATGGCCCAAG 5586
Qy 5281 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5340
Db 5587 GTAGAACAATTCTGGGCTAAACACATGTGGAACTTCATCAGCGGCATTCAATACCTCGCA 5646
Qy 5341 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5400
Db 5647 GGACTGTCAACACTGCCAGGGAACCCCGCTGTGGCTTCCATGATGGCATTCAGTGCCGCC 5706
Qy 5401 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5460
Db 5707 CTCACCAGTCCTCTGTCAACTAGCACCACTATCCTTCTCAACATTTTGGGAGGCTGGCTG 5766
Qy 5461 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5520
Db 5767 GCGTCCCAAATCGCGCCACCCGCAGGGGCCACCGGCTTCGTCGTCAGTGGCCTGGTGGGG 5826
Qy 5521 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5580
Db 5827 GCAGCTGTAGGCAGCATAGGCCTGGGCAAGGTGCTGGTGGACATCCTGGCAGGGTATGGT 5886
Qy 5581 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5640
Db 5887 GCGGGCATTTCGGGGGCCCTCGTCGCATTTAAGATCATGTCTGGCGAGAAGCCCTCCATG 5946
Qy 5641 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 5700
Db 5947 GAGGATGTTGTCAATCTGCTGCCTGGGATTTTGTCTCCGGGTGCCTTAGTGGTGGGAGTC 6006
Qy 5701 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 5760
Db 6007 ATCTGCGCGGCCATTCTACGCCGCCACGTGGGACCGGGGGAAGGTGCGGTCCAATGGATG 6066
Qy 5761 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 5820
Db 6067 AACAGGCTCATTGCCTTTGCTTCCAGAGGAAACCACGTCGCCCCTACCCACTACGTGACG 6126
Qy 5821 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 5880
Db 6127 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTGCTTGGATCCCTTACTATAACTAGTCTA 6186
Qy 5881 CTCAGAAGACTCCACAACTGGATTACTGAGGACTGCCCCATCCCATGCGGCGGCTCGTGG 5940
Db 6187 CTCAGGAGACTCCACAATTGGATTACTGAGGACTGCCCCATCCCATGCGCCGGTTCGTGG 6246
Qy 5941 CTCCGCGATGTGTGGGACTGGGTTTGCACCATCCTAACAGACTTTAAAAATTGGCTGACC 6000
Db 6247 CTCCGTGATGTGTGGGATTGGGTTTGCACCATCCTAACAGACTTTAAAAATTGGCTGACC 6306
Qy 6001 TCCAAATTATTCCCAAAGATGCCCGGCCTCCCCTTTGTCTCCTGTCAAAAGGGGTACAAG 6060
Db 6307 TCCAAATTGTTTCCAAAGATGCCCGGCCTCCCCTTTATCTCTTGCCAGAAGGGGTACAAG 6366
Qy 6061 GGCGTGTGGGCCGGCACTGGCATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6120
Db 6367 GGCGTGTGGGCCGGCACTGGTATCATGACCACACGATGTCCTTGCGGCGCCAACATCTCT 6426
Qy 6121 GGCAATGTCCGCTTGGGCTCCATGAGAATCACGGGGCCTAAGACCTGCATGAATATCTGG 6180
Db 6427 GGTAATGTCCGCCTGGGCTCTATGAGAATCACAGGGCCTAAAACCTGCATGAACACTTGG 6486
Qy 6181 CAGGGGACCTTTCCTATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCGCCA 6240
Db 6487 CAGGGGACCTTTCCTATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCGCCA 6546
Qy 6241 AACTTTAAGGTCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6300
Db 6547 AACTTCAAGGTCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6606
Qy 6301 GGGTCATACCACTACATAACAGGACTCACCACTGATAACTTGAAAGTCCCCTGCCAACTA 6360
Db 6607 GGATCATACTCCTATATAACAGGATTGACCACTGACAATTTGAAAGTTCCCTGCCAACTA 6666
Qy 6361 CCCTCTCCCGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCACA 6420
Db 6667 CCTTCTCCAGAGTTCTTCTCCTGGGTGGATGGGGTGCAAATCCATAGGTTCGCTCCCACA 6726
Qy 6421 CCGAAGCCGTTTTTCCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6480
Db 6727 CCAAAGCCGTTTTTCCGGGATGAGGTCTCGTTCTGCGTTGGACTCAATTCATTTGTTGTC 6786
Qy 6481 GGGTCCCAGCTTCCTTGCGACCCTGAACCCGACACAGACGTATTGATGTCCATGCTAACA 6540
Db 6787 GGGTCTCAGCTTCCTTGTGACCCTGAGCCCGACACTGACGTACTGACGTCCATGCTAACA 6846
Qy 6541 GATCCATCTCATATCACGGCGGAGACTGCAGCGCGGCGTTTAGCGCGGGGGGTCACCCCCA 6600
Db 6847 GATCCATCCCATATCACGGCGGAGACCGCAGCGCGGCGCTTGGCGCGGGGGTCACCTCCA 6906
Qy 6601 TCCGAGGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 6660
Db 6907 TCCGAGGCAAGCTCCTCAGCGAGCCAGCTATCAGCGCCATCGCTGCGAGCCACCTGCACC 6966
Qy 6661 ACCCACGGCAAAGCCTATGATGTGGACATGGTGGATGCTAACCTGTTCATGGGGGGCGAT 6720
Db 6967 ACCCATGGCAAGACTTATGATGTGGACATGGTGGATGCCAGCCTGTTCATGGGGGGCGAT 7026
Qy 6721 GTGACTCGGATAGAGTCTGGGTCCAAAGTGGTCGTTCTGGACTCTCTCGACCCAATGGTC 6780
Db 7027 GTGACTCGGATAGAGTCTGAGTCCAAAGTGGTCGTCCTGGACTCTCTCGACCCAATGGCC 7086
Qy 6781 GAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCAGAATACATGCTCCCCAAGAAGAGG 6840
Db 7087 GAAGAAGAGAGCGACCTCGAGCCCTCGATACCATCGGAGTATATGCTCTCCAGGAACAGG 7146
Qy 6841 TTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCGCTTGTGGAATCG 6900
Db 7147 TTCCCGCCAGCCTTACCGGCCTGGGCACGGCCTGATTACAACCCACCGCTTGTGGAGTCG 7206
Qy 6901 TGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTCCCTCCTCCTAGG 6960
Db 7207 TGGAAGAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTCCCCCCCCCTAAG 7266


Qy 6961 AAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGTGAGGACTCCATA 7020
Db 7267 AAGGCCCCGACGCCCCCCCCAAGGAGACGCCGGACAGTGGGTCTGAGCGAGAGCACCATA 7326
Qy 7021 GGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCCCCAAGCGGCGAT 7080
Db 7327 GGAGACGCCCTCCAACAGCTGGCCATCAAGTCCTTCGGCCAGCCCCCCCCAAGCGGCGAT 7386
Qy 7081 TCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACGCCTCCTGATGAG 7140
Db 7387 TCAGGCCTTTCCACGGGGGCGGACGCCGCCGACTCCGGCAGTCGGTCGCCCCCTGATGAG 7446
Qy 7141 TTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCTCGAGGGGGAGCTTGGA 7200
Db 7447 TTGGCTCTTTCGGAAACAGGTTCTACCTCCTCCATGCCCCCCCTCGAGGGGGAGCCTGGG 7506
Qy 7201 GATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCCAGGGGGGGGTGGCA 7260
Db 7507 GATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCAGGGGGGGGAGGTG 7566
Qy 7261 GCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGACGACTCCGTCGTG 7320
Db 7567 GCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGAATGTCTCCGTCGTT 7626
Qy 7321 TGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGTAGTCCCGAAGAG 7380
Db 7627 TTCTGCTCCATTTCATACTCCTGGACCGGGGGTCTAATAACTCCTTGTATCCCCGAAAAA 7686
Qy 7381 GAGAAGTTACCGATTAACCCCTTGAGCAACTCCCTGTTGCGATATCACAACAAGGTGTAC 7440
Db 7687 GAAAAGTTGCCATTTAACCCCTTAAGCAACTCCCTGTTGCGATACGACAACAAGGTGTAC 7746
Qy 7441 TGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTTGATAGGATGCAA 7500
Db 7747 TGTACTACATCAAAGAGCGCCTCACTGAGGGCTAAAAAGGTAACTTTTGATAGGATGCAA 7806
Qy 7501 GTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCGGCCTCCAAGGTC 7560
Db 7807 GTGCTCGACGCCTATTATGACTCAGTCTTAAAGGACATCAAGCTAGCGGCCTCCAAAGTC 7866
Qy 7561 ACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCCCATTCTGCAAGA 7620
Db 7867 AGCGCAAGGCTCCTCACCTTGGAGGAGGCGTGCCAATTGACTCCACCCCATTCCGCAAGA 7926
Qy 7621 TCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGGGCCGTTAACCAC 7680
Db 7927 TCCAAGTATGGGTTTGGGGCTAAGGAGGTTCGCAGCTTGTCCGGGAGGGCCGTTAACCAC 7986
Qy 7681 ATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATTCCCACAACCATT 7740
Db 7987 ATCAAGTCCGTGTGGAAGGACCTCCTGGAAGACTCACAAACACCAATTCCCACGACCATC 8046
Qy 7741 ATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAGAAAGCAGCTCGC 7800
Db 8047 ATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCAACAAGGGGGGTAAGAAAGCAGCTCGC 8106
Qy 7801 CTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCCCTTTATGACATT 7860
Db 8107 CTTATCGTTTACCCTGACCTCGGTGTCAGAGTTTGCGAGAAGATGGCCCTTTATGATGTC 8166
Qy 7861 ACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAGTATTCCCCCGCT 7920
Db 8167 ACACAAAAGCTTCCTCAGGCAGTGATGGGGGCCTCTTATGGCTTCCAGTACTCCCCCGCT 8226
Qy 7921 CAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCTATGGGTTTTTCG 7980
Db 8227 CAACGGGTGGAGTTTCTCTTGAAGGCATGGGCGGAAAAAAAAGACCCTATGGGGTTTTCG 8286
Qy 7981 TATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGGACTGAGGAGTCC 8040
Db 8287 TATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGAACTGAGGAGTCC 8346
Qy 8041 ATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACACTCGCTAACTGAG 8100
Db 8347 ATATACCGGCCTTGTTCCTTGCCCGAGGAGGCCCACACTGCCATACACTCGCTGACTGAG 8406
Qy 8101 AGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGCGGGTACAGGCGT 8160
Db 8407 AGACTTTACGTGGGAGGGCCCATGTTTAACAGCAAGGGCCAGGCCTGCGGGTACAGGCGT 8466
Qy 8161 TGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACATGCTACGTGAAA 8220
Db 8467 TGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACATGCTATGTGAAA 8526
Qy 8221 GCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTGGTATGCGGCGAT 8280
Db 8527 GCCCTAGCGGCCTGTAAGGCTGCGGGGATAGTTGCGCCCACAATGCTGGTATGCGGCGAC 8586
Qy 8281 GACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGGAACCTGAGAGCC 8340
Db 8587 GACTTGGTTGTCATCTCAGAAAGCCAGGGGACTGAGGAGGACGAGCGGAACCTGAGAGCC 8646
Qy 8341 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCCAGACCGGAGTAT 8400
Db 8647 TTCACGGAGGCTATGACCAGGTACTCTGCCCCTCCTGGTGACCCTCCCAGACCGGAATAT 8706
Qy 8401 GATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTGGGCCCACAAGGC 8460
Db 8707 GACCTGGAGCTGATAACATCTTGTTCCTCAAATGTGTCTGTGCACTTGGGCCCACAAGGC 8766
Qy 8461 CGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGGGCTGCCTGGGAA 8520
Db 8767 CGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGGGCTGCCTGGGAA 8826
Qy 8521 ACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAGTACGCCCCGACC 8580
Db 8827 ACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAGTACGCTCCAACC 8886
Qy 8581 ATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATGGCTCAAGACACG 8640
Db 8887 GTATGGGTTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATGGCCCAAGACACC 8946
Qy 8641 CTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTGAGTCCCTTGGAC 8700
Db 8947 CTGGACCAGAACCTCAACTTTGAAATGTACGGATCAGTGTACTCTGTGAGTCCTTTGGAC 9006
Qy 8701 CTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTGCACACATACACT 8760
Db 9007 CTCCCAGCCATAATTGAAAGGTTACACGGGCTTGACGCTTTCTCTCTGCACACATACTCT 9066
Qy 8761 CCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGA 8820
Db 9067 CCCCACGAACTGACGCGGGTGGCTGCAGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGA 9126
Qy 8821 GCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGTGGGGGGAGAGCG 8880
Db 9127 GCGTGGAAGAGTCGGGCGCGTGCAGTTAGGGCGTCCCTCATCTCCCATGGAGGGAGGGCG 9186
Qy 8881 GCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 8940
Db 9187 GCCGTTTGCGGTCGGTATCTCTTCAACTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 9246
Qy 8941 TTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGCGCCGGCGGGGGC 9000
Db 9247 TTGCCGGAAGCACGCCACCTGGATTTATCCAGTTGGTTTACTGTCGGCGCCGGCGGGGGC 9306
Qy 9001 GACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTTGGCCTACTCCTA 9060
Db 9307 CACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTACTCCTTGGCCTACTCCTA 9366
Qy 9061 CTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCGG 9099
Db 9367 CTTTTCGTAGGGGTAGGCCTCTTCCTACTCCCCGCTCGG 9405

RESULT 14

```
AF169002
LOCUS       AF169002                9661 bp    RNA     linear   VRL 06-SEP-2001
DEFINITION  Hepatitis C virus isolate NDM228 polyprotein gene, complete cds.
ACCESSION   AF169002
VERSION     AF169002.1  GI:6707279
KEYWORDS    .
SOURCE      Hepatitis C virus
  ORGANISM  Hepatitis C virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepacivirus.
REFERENCE   1  (bases 1 to 9661)
  AUTHORS   Kurihara,C., Ishiyama,N., Nishiyama,Y., Fukushi,S., Kageyama,T.,
            Katayama,K. and Miura,S.
  TITLE     Molecular characterization of hepatitis C virus genotype 2a from
            the entire sequences of four isolates
  JOURNAL   J. Med. Virol. 64 (4), 466-475 (2001)
  MEDLINE   21361470
   PUBMED   11468731
REFERENCE   2  (bases 1 to 9661)
  AUTHORS   Katayama,K. and Kurihara,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-1999) Basic Research Division, BioMedical
            Laboratories, Inc., 1361-1 Matoba, Kawagoe, Saitama 350-1101, Japan
FEATURES             Location/Qualifiers
     source          1. .9661
                     /organism="Hepatitis C virus"
                     /mol_type="genomic RNA"
                     /isolate="NDM228"
                     /db_xref="taxon:11103"
                     /note="genotype: 2a"
     CDS             341. .9442
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="AAF25610.1"
                     /db_xref="GI:6707280"
                     /translation="MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRL
                     GVRATRKTSERSQPRRRRQPIPKDRRSTGRSWGKPGYPWPLYGNEGLGWAGWLLSPRG
                     SRPSWGPNDPRHRSRNVGKVIDTLTCGFADLMGYIPVVGAPLGGVATALAHGVRALED
                     GINFATGNLPGCSFSIFLLALLSCITVPVSAAQVKNTSSGYMVTNDCANDSITWQLQA
                     VVLHVPGCVPCEXVXXTSRCWIPVSPNVAVRQPGALTQGLRTHLDMVVMSATLCSALY
                     VGDLCGGVMLAAQMFIVSPEHHWFVQECNCSIYPGAITGQRMAWDMMMNWSPTATMIL
                     AYAMRVPEVIIDIISGAHWGVMFGLAYFSMQGAWAKVVVILLLAAGVDANTXXXGGAA
                     ARNAYGVTSLFSXGAXQNIQLINTNGSWHINRTALNCNDSLNTGFLAXLFYXNRFNSS
                     GCPERLSVCRNIEAFRIGWGTLQYEDNVTNPEDMRPYCWHYPPKQCSIVPARSVCGPV
                     YCFTPSPVVVGTTDRRGVPTYTWGENETDVFLLNSTRPPQGSWFGCTWMNSTGFTKTC
                     GAPPCRTRADFNASTDLLCPTDCFRKHPEATYIKCGSGPWLTPRCLVDYPYRLWHYPC
                     TVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCDLEDRDRSQLSPLLHSTTEWAILPCS
                     YSDLPALSTGLLHLHQNIVDVQYMYGLSPALTKYIVRWEWVVLLFLLLADARVCACLW
                     MFILMGQAEAALEKLVVLHAASAASCNGFLYFVIFFVAAWYIKGRAVPLAAYSLTGLW
                     SFCLLLLVLPQQAYAYEASVHGQIGAALLVLITLFTLTPGYKTLLSRCLWWLCYLLTL
                     GEAMVQEWAPSMQARGGRDGIIWAATIFCPGVVFDITKWLLAALGSGYLLRGALTRVP
                     YFVRAHALLRMCTIAKHLAGGKYVQMALLALGRWTGTYIYDHLTPMSDWAASGLRDLA
                     VAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREILLGPADGYTSKGWRLL
                     APITAYAQQTRGLLGAIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISGVLWTVYHG
                     AGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRNADVIPA
                     RRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGIFRAAVCSRGVAKSIDFIPVET
                     LDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYASQGYKVLVLNPSV
                     AATLGFGAYLSKAHGINPNIRTGVRTVTTGEAITYSTYGKFLADGGCAGGAYDIIICD
                     ECHAVDATTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGHEGEIP
                     FYGKAIPLSYIKGGRHLIFCHSKKKCDELSAALRGMGLNAVAYYRGLDVSAIPAQGDV
                     VVVATDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQTVPQDAVSRSQRRGR
                     TGRGRLGIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPAETTVRLRAYFNTPGL
                     PVCQDHLEFWEAVFTGLTHIDAHFLSQTKQSGENFAYLVAYQATVCARAKAPPPSWDV
                     MWKCLTRLKPTLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAG
                     GVLAAVAAYCLATGCVSIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQ
                     RIAEMLKSKIQGLLQQASKQAQDIQPAVQASWPKVEQFWAKHMWNFISGIQYLAGLST
                     LPGNPAVASMMAFSAALTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAA
                     VGSVGLGKVLVDILAGYGAGISGALVAFKVMSGEKPSMEDVVNLLPGILSPGALVVGV
                     ICAAILRRHVGPGEGAVQWMNRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTIT
                     SLLRRLHNWITEDCPIPCAGSWLRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFISCQ
                     KGYRGVWAGTGIMTTRCSCGAEISGNVRLGSMRITGPRTCMNTWQGTFPINCYTEGQC
                     VPKPAPNFKTAIWRVAASEYAEVTQHGPYSYITGLTTDNLKVPCQIPSPEFFSWVDGV
                     QIHRFAPKPKPFFRDEVSFCVGLNSFVVGSQLPCDPEPDTDVLMSMLTDPSHITAEAA
                     ARRLARGSPPSEASSSASQLSAPSLRATCATHGKTYDVDMVDANLFMGGDVTRIESES
                     KVVVLDSLDPMAEERSDLEPSIPSEYMLPRNRFPPALPAWARPDYNPPLVESWKRPDY
                     QPPTVAGCALPPPKKAPTPPPRRRRTVGLSESTIGDALQQLAIKTFGQPPPSGDSGLS
                     TGADAADSGGQASPDELALSETGSTSSMPPLEGEPGDPDLEPERVELQALPQGGEVAP
                     GSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPSGPEEEKLPINPLSNSLLRYHNKVY
                     CTTSRTASLRAKKVTFDRTQVLDAYYDSTLKDIKLAASRVSARLLTLEEACQLTPPHS
                     ARSRYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSQTPIPTTIMAKNEVFCVEPTKGGK
                     KAARLIVYPDLSVRVCEKMALYDVTQKLPQAVMGASYGFQYSPAQRVEFLLKAWAEKK
                     DPMGFSYDTRCFDSTVTERDIRTEESIYQACSLPEEARTAIHSLTERLYVGGPMLNSK
                     GQACGYRRCRASGVLTTSMGNTITCYIKALAACKAAGIVAPTMLVCGDDLVVISESQG
                     TEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQGRRRYYLTR
                     DPTTPXARAAWETVRHSPVNSWLGNIIQYAPTIWVRMVLMTHFYSILMAQDTLDQNLN
                     FEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVAAALRKLGAPPLRAWKS
                     RARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARHLDLSSWFTVGAGGGDI
                     YHSVLRARPRLLLLGLLLLFVGVGLFLLPAR"
ORIGIN
```

Query Match 86.4%; Score 7858.6; DB 14; Length 9661;
Best Local Similarity 91.1%; Pred. No. 0;
Matches 8293; Conservative 41; Mismatches 765; Indels 0; Gaps 0;

```
Qy          1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 60
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db        341 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACTAACCGTCGCCCACAA 400

Qy         61 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 460

Qy        121 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        461 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTAGG 520

Qy        181 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 240
              |||||||||||||||| |||||||||||||||||||||  ||||||||||||||||||
Db        521 AGGCGCCAGCCCATCCCCAAAGATCGGCGCTCCACTGGCAGGTCCTGGGGAAAACCAGGA 580

Qy        241 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 300
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        581 TACCCCTGGCCCCTATATGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 640

Qy        301 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        641 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAATGTGGGT 700

Qy        361 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 420
              || |||||||| ||||||||||||||||||||||||||||||||||||||| |||||
Db        701 AAAGTCATCGACACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCCGTCGTA 760

Qy        421 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 480
              |||||||||||| ||:|||||||||| |||||||||||| |||||||||  |||||||||
Db        761 GGCGCCCCGCTTGGYGGCGTCGCCACAGCTCTCGCGCACGGCGTGAGAGCTCTGGAGGAC 820

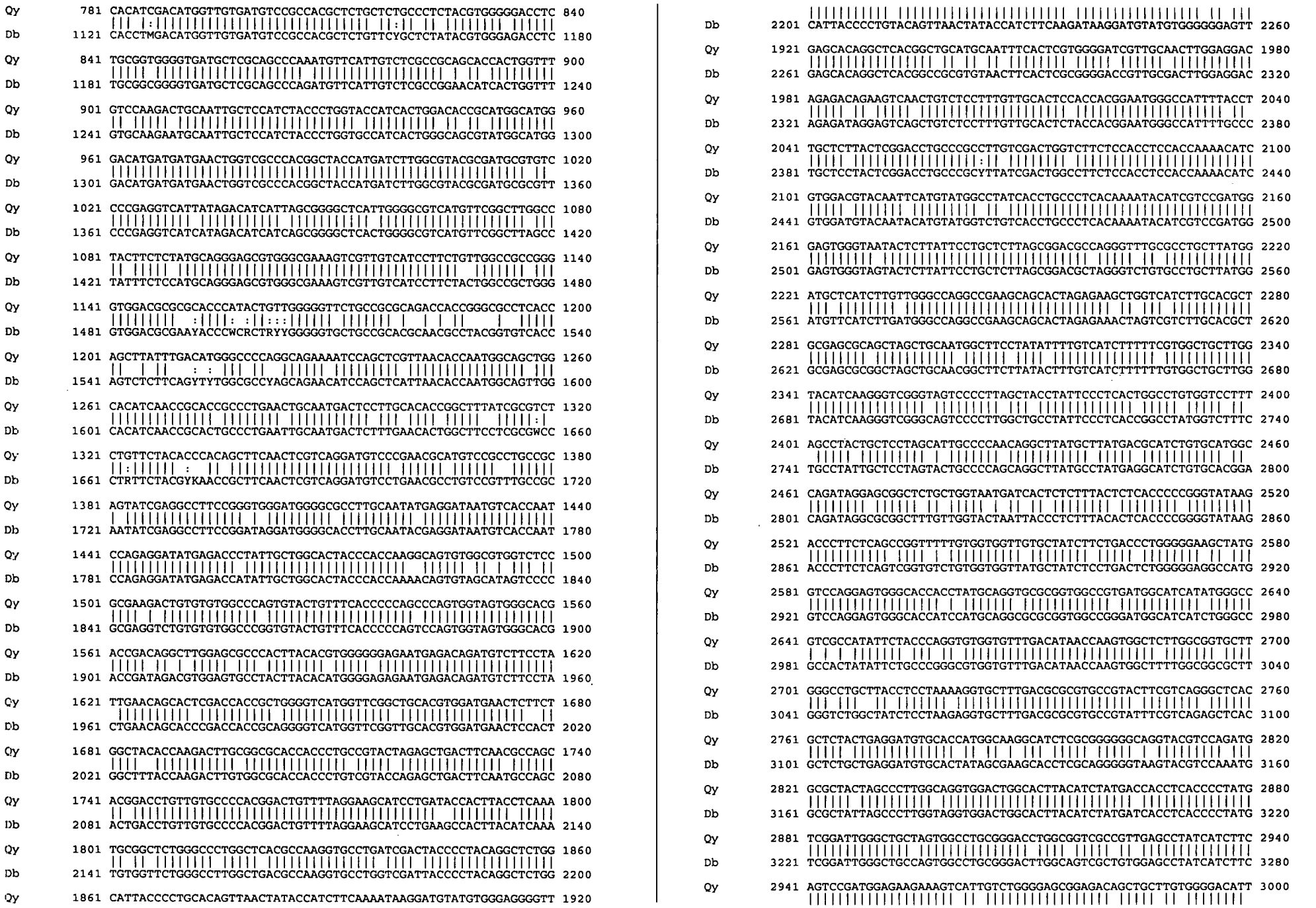
Qy        481 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 540
              ||| |||||||||||||||||:|||| ||||||||||||||||||||||||||||||||||
Db        821 GGGATTAATTTTGCAACAGGGAACYTACCTGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 880

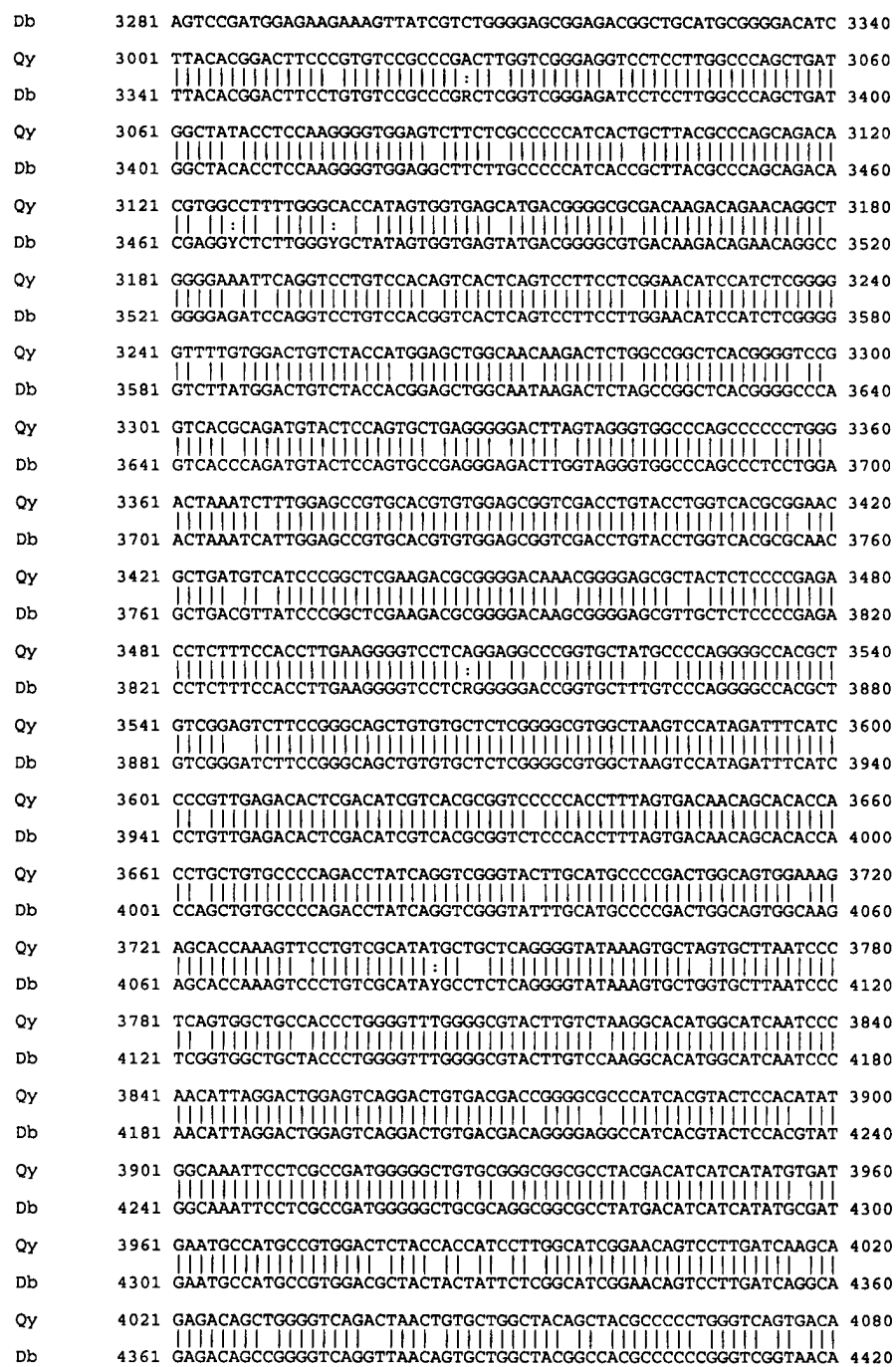
Qy        541 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 600
              ||| |||| |||||||| ||||||||||||||| |||||||||||  ||:| |||
Db        881 CTGTTGTCTTGCATCACTGTTCCGGTCTCCGCTGCCCAAGTGAAGAACACTAGYAGCGGT 940

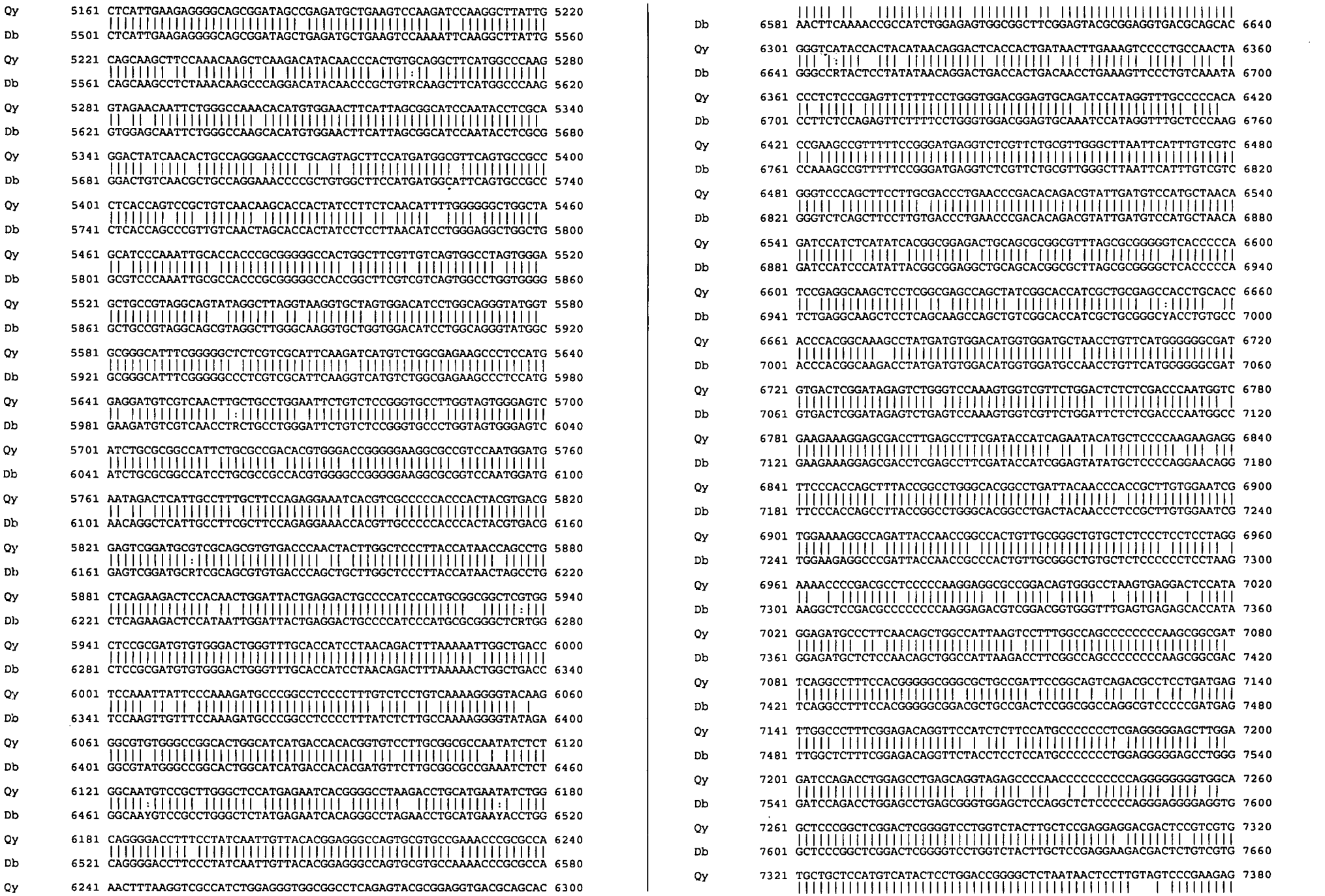
Qy        601 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 660
              |||||||||| |||||||| |||||||||||||| |||||||||| ||||| | |||
Db        941 TACATGGTGACCAACGACTGTGCCAATGACAGCATCACCTGGCAGCTTCAGGCCGTGGTC 1000

Qy        661 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 720
              |||||||||| || || ||||| ||||||:: |||:: ::  | || | ||| ||||||
Db       1001 CTCCACGTCCCTGGATGTGTCCCATGCGAGMRCGTGRRTRRCACCTCGCGGTGTTGGATA 1060

Qy        721 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 780
              |||||||||| || ||||| |||| || |||||||||||||||||| ||| ||||||||
Db       1061 CCGGTCTCACCAAACGTGGCTGTGCGGCAGCCCGGCGCCCTCACGCAAGGCCTGCGGACG 1120
```




Db 3281 AGTCCGATGGAGAAGAAAGTTATCGTCTGGGGAGCGGAGACGGCTGCATGCGGGGACATC 3340
Qy 3001 TTACACGGACTTCCCGTGTCCGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3060
Db 3341 TTACACGGACTTCCTGTGTCCGCCCGRCTCGGTCGGGAGATCCTCCTTGGCCCAGCTGAT 3400
Qy 3061 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3120
Db 3401 GGCTACACCTCCAAGGGGTGGAGGCTTCTTGCCCCCATCACCGCTTACGCCCAGCAGACA 3460
Qy 3121 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCT 3180
Db 3461 CGAGGYCTCTTGGGYGCTATAGTGGTGAGTATGACGGGGCGTGACAAGACAGAACAGGCC 3520
Qy 3181 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3240
Db 3521 GGGGAGATCCAGGTCCTGTCCACGGTCACTCAGTCCTTCCTTGGAACATCCATCTCGGGG 3580
Qy 3241 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3300
Db 3581 GTCTTATGGACTGTCTACCACGGAGCTGGCAATAAGACTCTAGCCGGCTCACGGGGCCCA 3640
Qy 3301 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGG 3360
Db 3641 GTCACCCAGATGTACTCCAGTGCCGAGGGAGACTTGGTAGGGTGGCCCAGCCCTCCTGGA 3700
Qy 3361 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3420
Db 3701 ACTAAATCATTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGCAAC 3760
Qy 3421 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3480
Db 3761 GCTGACGTTATCCCGGCTCGAAGACGCGGGGACAAGCGGGGAGCGTTGCTCTCCCCGAGA 3820
Qy 3481 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3540
Db 3821 CCTCTTTCCACCTTGAAGGGGTCCTCRGGGGGACCGGTGCTTTGTCCCAGGGGCCACGCT 3880
Qy 3541 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3600
Db 3881 GTCGGGATCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3940
Qy 3601 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 3660
Db 3941 CCTGTTGAGACACTCGACATCGTCACGCGGTCTCCCACCTTTAGTGACAACAGCACACCA 4000
Qy 3661 CCTGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 3720
Db 4001 CCAGCTGTGCCCCAGACCTATCAGGTCGGGTATTTGCATGCCCCGACTGGCAGTGGCAAG 4060
Qy 3721 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 3780
Db 4061 AGCACCAAAGTCCCTGTCGCATAYGCCTCTCAGGGGTATAAAGTGCTGGTGCTTAATCCC 4120
Qy 3781 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 3840
Db 4121 TCGGTGGCTGCTACCCTGGGGTTTGGGGCGTACTTGTCCAAGGCACATGGCATCAATCCC 4180
Qy 3841 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 3900
Db 4181 AACATTAGGACTGGAGTCAGGACTGTGACGACAGGGGAGGCCATCACGTACTCCACGTAT 4240
Qy 3901 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 3960
Db 4241 GGCAAATTCCTCGCCGATGGGGGCTGCGCAGGCGGCGCCTATGACATCATCATATGCGAT 4300
Qy 3961 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4020
Db 4301 GAATGCCATGCCGTGGACGCTACTACTATTCTCGGCATCGGAACAGTCCTTGATCAGGCA 4360
Qy 4021 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4080
Db 4361 GAGACAGCCGGGGTCAGGTTAACAGTGCTGGCTACGGCCACGCCCCCCGGGTCGGTAACA 4420
Qy 4081 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4140
Db 4421 ACCCCCCATCCCAATATAGAGGAGGTAGCCCTAGGGCACGAGGGTGAGATCCCCTTCTAT 4480
Qy 4141 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4200
Db 4481 GGGAAGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGGCACTTGATCTTCTGTCACTCA 4540
Qy 4201 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4260
Db 4541 AAGAAAAAGTGTGACGAGCTCTCGGCGGCCCTTAGGGGCATGGGCTTGAACGCTGTGGCA 4600
Qy 4261 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4320
Db 4601 TATTACAGAGGGTTGGACGTCTCCGCAATACCAGCTCAGGGAGATGTAGTGGTCGTCGCC 4660
Qy 4321 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4380
Db 4661 ACCGACGCCCTCATGACGGGGTATACTGGAGACTTTGACTCCGTGATCGACTGCAAYGTA 4720
Qy 4381 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4440
Db 4721 GCGGTCACTCAAGTCGTAGACTTCAGCTTAGACCCCACCTTCACCATAACCACACAGACC 4780
Qy 4441 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4500
Db 4781 GTCCCTCAAGACGCTGTCTCACGCAGCCAGCGCCGGGGGCGTACAGGCAGAGGAAGACTG 4840
Qy 4501 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4560
Db 4841 GGCATCTATAGGTACGTTTCCACTGGTGAGCGAGCTTCAGGAATGTTTGACAGCGTAGTG 4900
Qy 4561 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4620
Db 4901 CTCTGTGAGTGCTACGACGCAGGGGCTGCRTGGTATGAGCTCACACCAGCGGAGACTACT 4960
Qy 4621 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 4680
Db 4961 GTCAGGCTGAGAGCATATTTCAACACGCCTGGCCTGCCCGTATGCCAAGATCACCTTGAG 5020
Qy 4681 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 4740
Db 5021 TTTTGGGAGGCAGTCTTCACCGGCCTCACACACATAGATGCCCACTTCCTGTCCCAAACA 5080
Qy 4741 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 4800
Db 5081 AAGCAATCGGGGGAGAATTTCGCATACTTAGTAGCCTACCAGGCTACAGTGTGCGCCAGA 5140
Qy 4801 GCCAAAGCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 4860
Db 5141 GCCAAAGCTCCTCCTCCGTCCTGGGACGTCATGTGGAAGTGCTTGACCCGACTCAAGCCC 5200
Qy 4861 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 4920
Db 5201 ACGCTTGTGGGCCCTACACCTCTCCTGTACCGTTTGGGCTCTGTTACCAACGAAGTCACC 5260
Qy 4921 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 4980
Db 5261 CTCACACATCCTGTGACGAAGTACATCGCCACATGTATGCAAGCTGATCTTGAGGTCATG 5320
Qy 4981 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5040
Db 5321 ACCAGCACGTGGGTCCTAGCTGGGGGAGTCCTGGCAGCCGTCGCMGCGTATTGTCTCGCG 5380
Qy 5041 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5100
Db 5381 ACCGGGTGTGTTTCCATCATCGGCCGCTTGCACATCAACCAGCGAGCCGTCGTTGCGCCG 5440
Qy 5101 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5160
Db 5441 GACAAGGAAGTCCTCTATGAGGCTTTTGAYGAGATGGAGGAATGTGCCTCTAGAGCGGCT 5500



```
Db       7661 TGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTCCGGCCCCGAAGAG 7720

Qy       7381 GAGAAGTTACCGATTAACCCCTTGAGCAACTCCCTGTTGCGATATCACAACAAGGTGTAC 7440
             ||||||| || || |||||||||||||||| |||||||||| |||||||||||||||||
Db       7721 GAGAAGTTGCCAATCAACCCCTTGAGCAACTCGCTGTTGCGATACCACAACAAGGTGTAC 7780

Qy       7441 TGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTTGATAGGATGCAA 7500
             ||||| ||| ||| || |||||||||| ||||||||||||||||||||||||||||| ||||
Db       7781 TGTACTACATCAAGGACCGCCTCACTGAGGGCTAAAAAGGTAACTTTTGATAGGACGCAA 7840

Qy       7501 GTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCGGCCTCCAAGGTC 7560
             ||||||||| |||| || |||||| |||||||||||| |||||||||||||||| ||||
Db       7841 GTGCTCGACGCCTATTATGACTCAACCTTAAAGGACATCAAGCTAGCGGCCTCCAGGGTC 7900

Qy       7561 ACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCCCATTCTGCAAGA 7620
             | |||||||||||| ||| |||||||||| || ||| |||| || ||||| |||||||||
Db       7901 AGCGCAAGGCTCCTTACCTTGGAGGAGGCGTGTCAGCTAACTCCGCCCCACTCTGCAAGA 7960

Qy       7621 TCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGGGCCGTTAACCAC 7680
             || |  || ||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       7961 TCCAGGTACGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGGGCCGTCAACCAC 8020

Qy       7681 ATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATTCCCACAACCATT 7740
             |||||||||||||||||||||||||||| |||||| |||||||||||||||| |||||
Db       8021 ATCAAGTCCGTGTGGAAGGACCTCCTGGAAGACTCACAAACACCAATTCCCACGACCATC 8080

Qy       7741 ATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAGAAAGCAGCTCGC 7800
             |||||||||||||||||||||||||||| |||||||||||| ||||||||||||||||||
Db       8081 ATGGCCAAAAATGAGGTGTTCTGCGTGGAACCCACCAAGGGGGGTAAGAAAGCAGCTCGC 8140

Qy       7801 CTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCCCTTTATGACATT 7860
             || |||||||||||||||| |||||||| |||||||||||||||||||||||  |
Db       8141 CTCATCGTTTACCCTGACCTCAGCGTCAGGGTTTGCGAGAAGATGGCCCTTTATGATGTC 8200

Qy       7861 ACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAGTATTCCCCCGCT 7920
             |||||||| |||||||||||||||||||||||||||||||||||||||| ||||||||
Db       8201 ACACAAAAGCTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAGTACTCCCCCGCT 8260

Qy       7921 CAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCTATGGGTTTTTCG 7980
             |||||||| |||||||| |||||||||||||||||||||||| |||||||||| ||||||
Db       8261 CAGCGGGTGGAGTTTCTTTTGAAAGCATGGGCGGAAAAGAAAGACCCTATGGGGTTTTCG 8320

Qy       7981 TATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGGACTGAGGAGTCC 8040
             |||||||||||||||||||||||| |||||||||||||||||| ||||||||||||
Db       8321 TATGATACCCGATGCTTTGACTCAACAGTCACTGAGAGAGACATCAGAACTGAGGAGTCC 8380

Qy       8041 ATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACACTCGCTAACTGAG 8100
             ||||| | |||||||||| |||||||||||||| |||||||||||||||| ||||||
Db       8381 ATATACCAGGCCTGCTCCCTGCCCGAGGAGGCCCGCACTGCCATACACTCGCTGACTGAG 8440

Qy       8101 AGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGCGGGTACAGGCGT 8160
             |||||||||| || ||||| ||| | || |||||||||| ||||||||||||||||||
Db       8441 AGACTTTACGTAGGTGGGCCCATGCTGAATAGCAAGGGCCAAGCCTGCGGGTACAGGCGT 8500

Qy       8161 TGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACATGCTACGTGAAA 8220
             |||||||||||||||||||||||||||||||||||||||||||||||||||| | |||
Db       8501 TGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACATGCTACATAAAA 8560

Qy       8221 GCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTGGTATGCGGCGAT 8280
             ||| |||||||||| || ||||| |||||| | ||||||||||||||||||||||||||
Db       8561 GCCCTAGCGGCTTGCAAGGCTGCGGGGATAGTTGCGCCCACAATGCTGGTATGCGGCGAC 8620

Qy       8281 GACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGGAACCTGAGAGCC 8340
             |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8621 GACTTGGTCGTCATCTCAGAAAGCCAGGGGACTGAGGAGGACGAGCGGAACCTGAGAGCC 8680

Qy       8341 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCCAGACCGGAGTAT 8400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db       8681 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCCAGACCGGAATAT 8740

Qy       8401 GATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTGGGCCCACAAGGC 8460
             || |||||||| |||||||||| |||||||||||||||||||||||||||| || |||
Db       8741 GACCTGGAGCTAATAACATCTTGTTCCTCAAATGTGTCTGTGGCGCTGGGCCCGCAGGGC 8800

Qy       8461 CGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGGGCTGCCTGGGAA 8520
             |||||||||||||:|||||||||||||||||||||| |||||||||||||||||||||
Db       8801 CGCCGCAGATACTAYCTGACCAGAGACCCTACCACTCCAYTCGCCCGGGCTGCCTGGGAA 8860

Qy       8521 ACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAGTACGCCCCGACC 8580
             |||||||||||||||| |||||||||||||||||||||||||||||||||||||| |||
Db       8861 ACAGTTAGACACTCCCCCGTCAATTCATGGCTGGGAAACATCATCCAGTACGCCCCAACC 8920

Qy       8581 ATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATGGCTCAAGACACG 8640
             |||||||| |||||||||||||||||||||||| | |||||||||||||| ||||||||
Db       8921 ATATGGGTTCGCATGGTCCTGATGACACACTTTTACTCCATTCTCATGGCCCAAGACACC 8980

Qy       8641 CTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTGAGTCCCTTGGAC 8700
             |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       8981 CTGGACCAGAACCTTAACTTTGAGATGTACGGAGCGGTGTACTCCGTGAGTCCCTTGGAC 9040

Qy       8701 CTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTGCACACATACACT 8760
             |||||||| |||||||||||||||| |||||||||||||||| |||||||||| |||
Db       9041 CTCCCAGCCATAATTGAAAGGTTACACGGGCTTGACGCTTTTTCACTGCACACATATACT 9100

Qy       8761 CCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGA 8820
             |||||||||||||| ||||||||| |||||||||||||||||||| ||||||||||||
Db       9101 CCCCACGAACTGACGCGGGTGGCTGCAGCCCTCAGAAAACTTGGGGCACCACCCCTCAGA 9160

Qy       8821 GCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGTGGGGGGAGAGCG 8880
             |||||||||| ||||| ||||| || |||||||||||||||||||||| |||||||||
Db       9161 GCGTGGAAGAGTCGGGCGCGTGCGGTTAGGGCGTCCCTCATCTCCCGTGGAGGGAGAGCG 9220

Qy       8881 GCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 8940
             ||||| |||||||| |||||||||| |||||||||||||||||||||||||||||||||
Db       9221 GCCGTCTGCGGTCGGTATCTCTTCAACTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 9280

Qy       8941 TTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGCGCCGGCGGGGGC 9000
             |||||||| || |||| |||||||||||||| |||||||||||||||||||||||||||
Db       9281 TTGCCGGAGGCACGCCACCTGGATTTATCCAGTTGGTTCACCGTCGGCGCCGGCGGGGGC 9340

Qy       9001 GACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTTGGCCTACTCCTA 9060
             |||||||||||||||||| ||| |||||||||||||||| ||| ||||||||||||||
Db       9341 GACATTTATCACAGCGTGTTGCGCGCCCGACCCCGCTTATTACTCCTTGGCCTACTCCTA 9400

Qy       9061 CTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCGG 9099
             ||||| |||||||||||||| ||||||||||||||||||
Db       9401 CTTTTCGTAGGGGTAGGCCTCTTCCTACTCCCCGCTCGG 9439
```

```
RESULT 15
AF238483
LOCUS       AF238483               9416 bp    RNA     linear   VRL 26-MAR-2000
DEFINITION  Hepatitis C virus 2a polyprotein gene, complete cds.
ACCESSION   AF238483
VERSION     AF238483.1  GI:7329204
KEYWORDS    .
SOURCE      Hepatitis C virus
  ORGANISM  Hepatitis C virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepacivirus.
REFERENCE   1  (bases 1 to 9416)
  AUTHORS   Itakura,J., Nagayama,K., Enomoto,N., Kurosaki,M., Watanabe,H. and
            Sato,C.
  TITLE     Full length cDNA sequence of HCV genotype 2a, strain MD2a-4
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 9416)
  AUTHORS   Itakura,J., Nagayama,K., Enomoto,N., Kurosaki,M., Watanabe,H. and
            Sato,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-FEB-2000) Second Department of Internal Medicine,
            Tokyo Medical and Dental University, 1-5-45 Yushima Bunkyo-Ku,
            Tokyo 113-8519, Japan
FEATURES             Location/Qualifiers
     source          1. .9416
                     /organism="Hepatitis C virus"
                     /mol_type="genomic RNA"
```

```
                     /strain="MD2a-4"
                     /db_xref="taxon:11103"
                     /note="genotype: 2a"
     CDS             307. .9408
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="AAF59942.1"
                     /db_xref="GI:7329205"
                     /translation="MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRL
                     GVRATRKTSERSQPRGRRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRG
                     SRPSWGPTDPRHRSRNVGKVIDTLTCGFADLMGYIPVVGAPLGGVARTLAHGVRVLED
                     GVNYATGNLPGCSFSIFLLALLSCITIPVSAIEVKNTSDGYMVTNDCSNGSITWQLQA
                     AVLHVPGCVPCEKVGNTSRCWIPVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALY
                     VGDLCGGVMLAAHMFIVTPQHHWFVQDCNCSIYPGTITGHRMAWDMMMNWSPTATMIL
                     AYAMRVPEVIIDIIGGAHWGVMFGLAYFSMQGAWAKVVVILLLTAGVDANTYSVGGST
                     AHLTQTFAGLFTLGSQQNIQLINTNGSWHINRTALNCNESLNTGFLASLFYTHAFNSS
                     GCPERLSACRNIEAFRIGWGVLQYEENVTNPEDMRPYCWHYPPKQCGIVPARSVCGPV
                     YCFTPSPVVVGTTDKLGVPTYTWGENETDVFLLNSTRPPRGAWFGCTWMNSTGFTKTC
                     GAPPCRIRADFNASTDLLCPTDCFRKHPEATYTKCGSGPWLTPRCMVDYPYRLGIHPC
                     TVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLEDRDRSQLSPLLHSTTEWAILPCT
                     FSDLPALSTGLLHLHQNIVDVQYMYGLSPAITRHIVRWEWVVLLFLLLADARVCACVW
                     MLILLGQAEAALEKLVVLHAASAASCNGFLYFVIFFVAAWYIKGRAVPLAAYSLTGLW
                     SFCLLLLALPQQAYAYDASVHGQIGAVLLILVTLFTLTPGYKILLSRCLWWLCYLLTL
                     GEAMIQEWVPPMQVRGGRDGIIWAATIFCPGVVFDITKWLLALLGPGYLLRGALTRVP
                     YFVRAHALLRMCTMARHLAGGKYVQMALLALGRWTGTYIYDHLTPMSDWAASGLRDLA
                     VAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPADGYTSKGWKLL
                     APITAYAQQTRGLLGAIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISGVLWTVYHG
                     AGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRNADVIPA
                     RRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGIFRAAVCSRGVAKSIDFIPVET
                     LDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLILNPSV
                     GATLGFGAYLSKAYGINPNIRTGVRTVTTGDSITYSTYGKFLADGGCAGGAYDIIICD
                     ECHAVGSTTNLGIGTVLDQAETAGVRLTVLATATLPGSVETPHPNIEEVALGQEGEIP
                     FYGRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNAVAYYRGLDVSVIPTQGDV
                     VVVATDALMTGFTGDFDTVIDCNVAVTQVVDFSLDPTFTITTQTVPQDAVSRSQRRGR
                     TGRGRLGTYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPAETTVRLRAYFNTPGL
                     PVCQDHLEFWEAVFTGLTHIDAHFLSQTKQAGESFPYLVAYQATVCARAKAPPPSWDI
                     MWKCLTRLKPTLVGPTPLLYRLGSVTNEITLTHPVTKYIATCMQADLEVMTSTWVLAG
                     GVLAAVAAYCLATGCVSIIGRLHINQRAIVAPDKEVLYEAFDEMEECASKAALLEEGQ
                     RIAEMLKSKIQGLLQQASKQAQDIQPAVQASWPKVEQFWAKHMWNFISGIQYLARLST
                     LPGNPAVASMMAFSAALTSPLSTSTTILLNILGRWLASQIAPPAGATGFVVSGLVGAA
                     VGNIGLGKVLVDILAGYGAGISGALVAFKIMSGEKPSMEDVINLLPGILSPGALVVGV
                     ICAAILRRHVGPGEGAVQWMNRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTIT
                     SLLRRLHNWITEDCPIPCAGSWLRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFISCQ
                     KGYKGVWAGTGIMTTRCPCGANISGNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQC
                     VPKPAPNFKTAIWRVAASEYAEVTQHGQYSYVTGLTTDNLKVPCQLPSPEFFSWVDGV
                     QIHRFAPTPKPFFRDEVSFYVGLNSFVVGSQLPCDPEPDTDVLMSMLTEPSHITAEAA
                     ARRLARGSPPSEASSSASQLSAPSLRATCTTHGKAYDVDMVDANLFMGGDVTRIESES
                     KVVVLDSLDPMAEEKSDLEPSIPSEYMLPRNRFPPALPAWARPDYNPPLVESWKRPDY
                     QPPTVAGCALPPPKKTPTPPPRRRRTVGLDESTIGDALQQLAIKTFGQIPPSADSGLS
                     TGADAADSGGRTSPDELALSETGSISSMPPLEGEPGDPDLEPEQVELQPPPQEGELAP
                     GSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEEEKLPINPLSNSLLRYHNKVY
                     CTTSKSASLRAKKVTFDRMQVLDAHYDSVLKDIKLAASKVSARLLTLEEACQLTPPHS
                     ARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSQTPIPTTIMAKNEVFCVDPTKGGK
                     KAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPAQRVEFLLKAWAEKK
                     DPLGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTERLYVGGPMLNSK
                     GQSCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIVAPTMLVCGDDLVVISESQG
                     TEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALNPLGRRRYYLTR
                     DPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDTLDQNLN
                     FEMYGSVYSVSPLDLPAIIERLHGLDAFSLHTYSPHELTRVASTLRKLGAPPLRAWKS
                     RARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGGDI
                     YHSVSRARPRLLLLGLLLLSVGVGLFLLPAR"
ORIGIN

  Query Match           86.3%;  Score 7852.2;  DB 14;  Length 9416;
  Best Local Similarity  91.4%;  Pred. No. 0;
  Matches 8319;  Conservative    0;  Mismatches  780;  Indels    0;  Gaps    0;

Qy          1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        307 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACTAACCGTCGCCCACAA 366

Qy         61 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        367 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 426

Qy        121 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        427 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGG 486

Qy        181 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 240
              ||||||||||||||||| ||||||||||||||||||||| ||||||||||||||||||||

Db        487 AGGCGCCAGCCCATCCCCAAAGATCGGCGCTCCACTGGCAAGTCCTGGGGAAAACCAGGA 546

Qy        241 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 300
              ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||

Db        547 TACCCCTGGCCCCTATATGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 606

Qy        301 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 360
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||

Db        607 CGAGGTTCCCGTCCCTCTTGGGGCCCCACTGACCCCCGGCATAGGTCGCGCAACGTGGGT 666

Qy        361 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||

Db        667 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCCGTCGTA 726

Qy        421 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 480
              |||||||||| || |||||||||| |||||||||| ||||||||||||||||||||||

Db        727 GGCGCCCCGCTTGGTGGCGTCGCCAGAACTCTCGCGCACGGCGTGAGAGTCCTGGAGGAC 786

Qy        481 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 540
              |||||||||| |||||||||||||||| ||||||||||||||||||||||||||||||||

Db        787 GGGGTTAATTATGCAACAGGGAACTTACCTGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 846

Qy        541 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 600
              || |||||||||||||||| |||||||| ||| |||||||||||||| |||| |||||||

Db        847 CTACTGTCCTGCATCACCATTCCGGTCTCTGCTATCGAAGTGAAGAACACCAGTGACGGT 906

Qy        601 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 660
              || |||||||||| ||||| |||||| |||||| |||||||| |||||||| || |||||

Db        907 TATATGGTGACTAATGACTGTTCCAATGGCAGCATCACCTGGCAACTCCAGGCCGCGGTC 966

Qy        661 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 720
              |||||||||||||||||||||||||||||||||||||||||||| ||| |||| | |||||||||

Db        967 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGAAATACATCACGGTGCTGGATA 1026

Qy        721 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 780
              |||||||||| || |||||||||||||||||||||||||||||||||||||| ||||||||

Db       1027 CCGGTCTCACCAAACGTAGCCGTGCAGCGGCCTGGCGCCCTCACGCAGGGCCTGCGGACG 1086

Qy        781 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 840
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||

Db       1087 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCTCTCTACGTGGGGGACCTC 1146

Qy        841 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 900
              ||||| |||||||||||||||||| |||||||||||| ||||||| ||||||||||||

Db       1147 TGCGGCGGGGTGATGCTCGCAGCCCATATGTTCATTGTCACGCCGCAACACCACTGGTTT 1206

Qy        901 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 960
              || || || |||||||||||||||||||||||||||||||||||||||||||||||||||

Db       1207 GTGCAGGATTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 1266

Qy        961 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1020
              |||||||||||||||||||||||||||||| |||||||| |||||||||||||||| |||

Db       1267 GACATGATGATGAACTGGTCGCCCACGGCCACCATGATCCTGGCGTACGCGATGCGCGTC 1326

Qy       1021 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1080
              |||||||||| |||||||||| | |||||||||||||||||||||||||||| |||

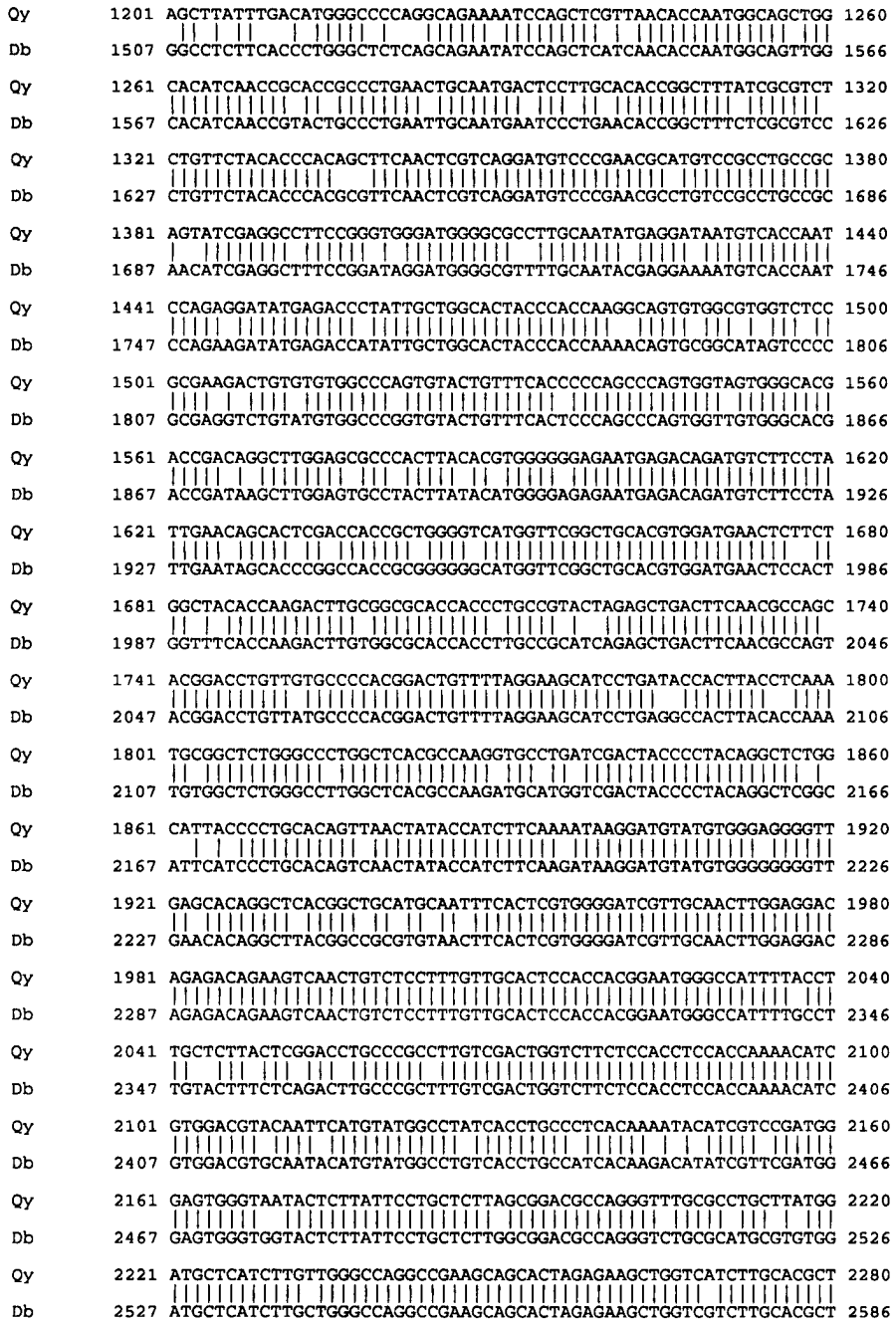
Db       1327 CCCGAGGTCATCATAGACATCATTGGTGGGGCTCATTGGGGCGTCATGTTCGGCTTAGCC 1386

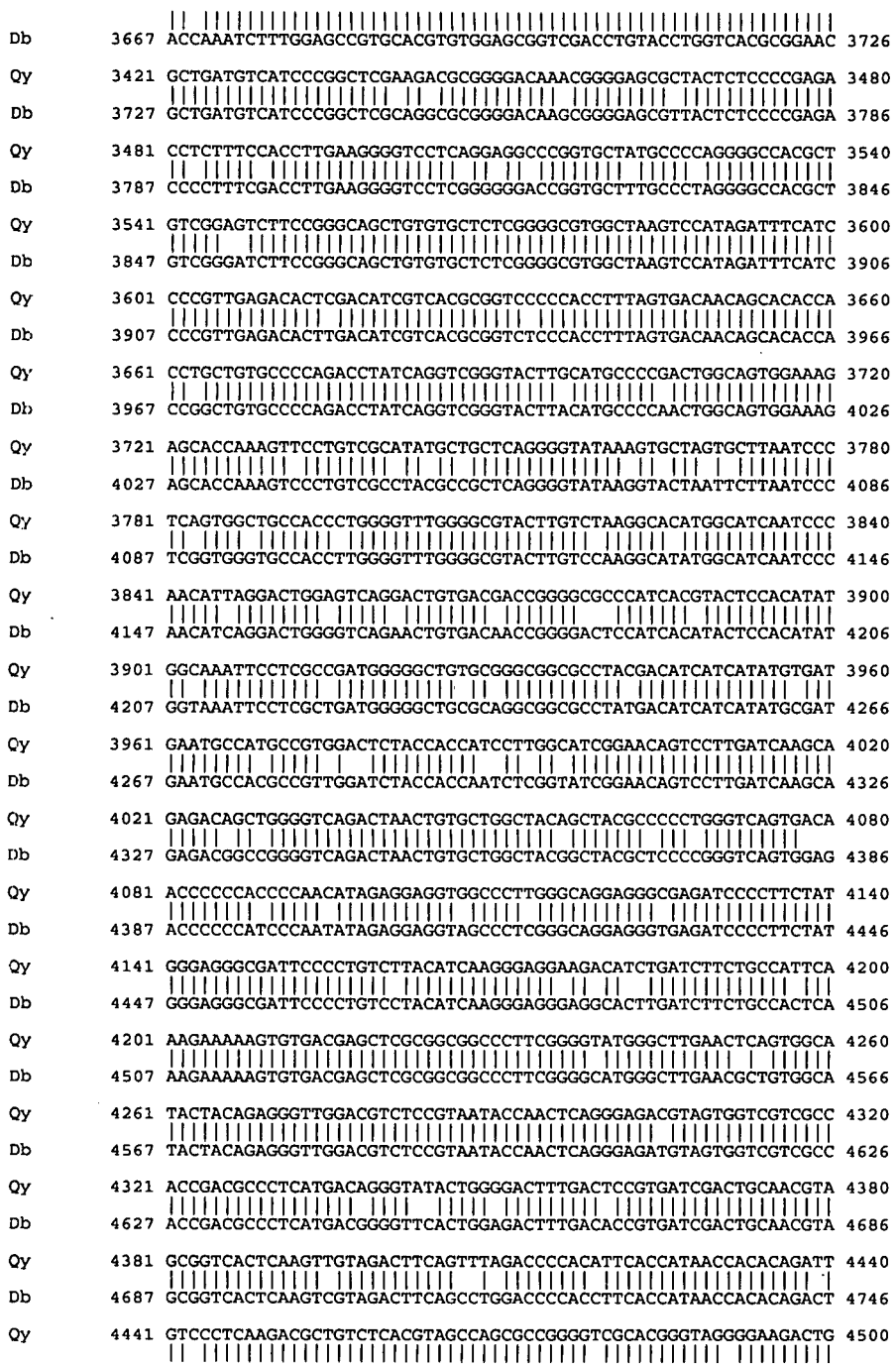
Qy       1081 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1140
              |||||||||||||||||||||||| |||| |||||||||||||| ||| || |||| |||

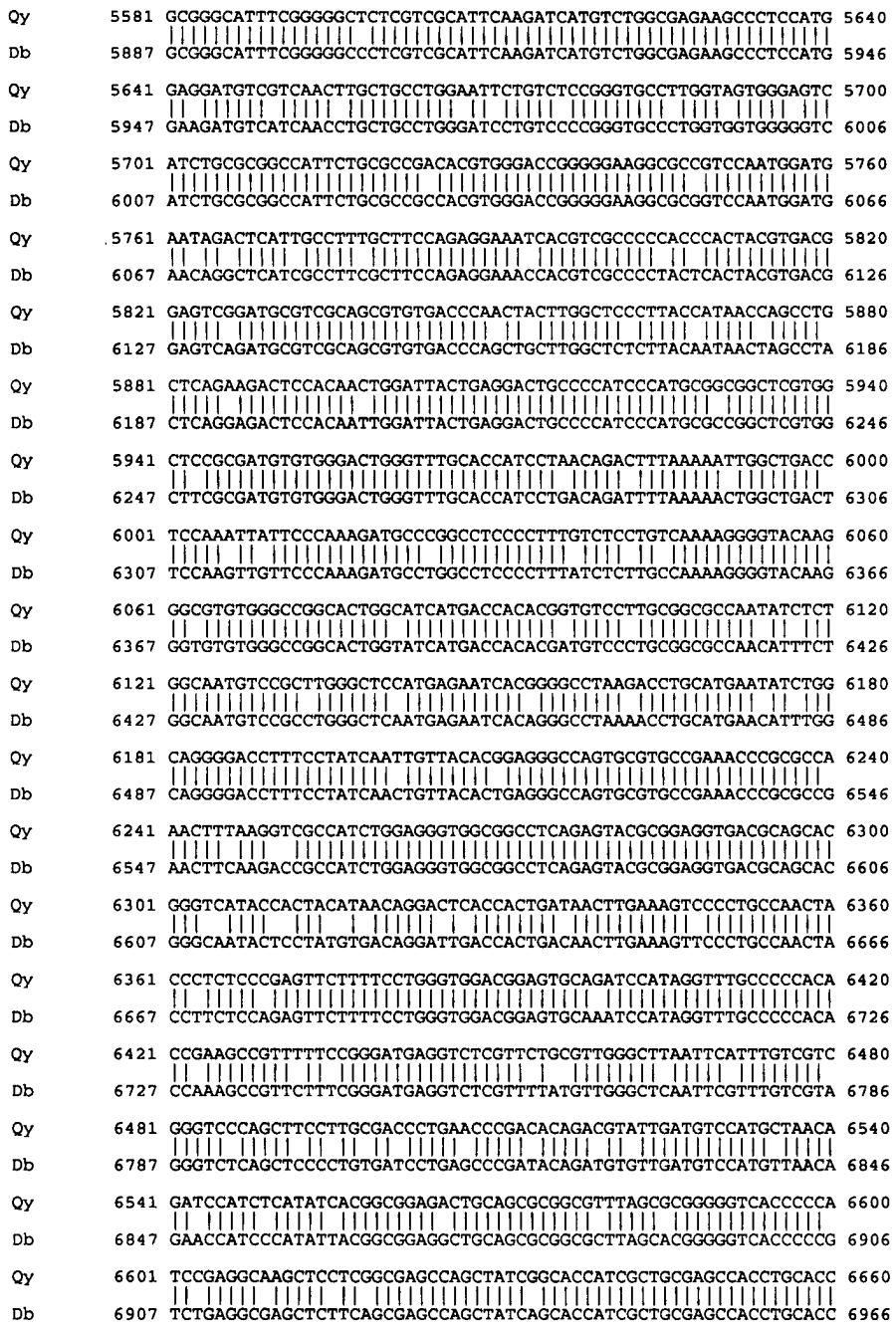
Db       1387 TACTTCTCTATGCAGGGAGCCTGGGCAAAAGTCGTTGTCATCCTCCTGCTGACCGCTGGG 1446

Qy       1141 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1200
              ||||||||   ||| | | || |||||||| | |||||  |||| | | | ||

Db       1447 GTGGACGCGAATACCTACTCCGTCGGGGGTTCCACTGCGCATCTCACCCAGACCTTTGCC 1506
```


Qy 1201 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1260
Db 1507 GGCCTCTTCACCCTGGGCTCTCAGCAGAATATCCAGCTCATCAACACCAATGGCAGTTGG 1566
Qy 1261 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1320
Db 1567 CACATCAACCGTACTGCCCTGAATTGCAATGAATCCCTGAACACCGGCTTTCTCGCGTCC 1626
Qy 1321 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1380
Db 1627 CTGTTCTACACCCACGCGTTCAACTCGTCAGGATGTCCCGAACGCCTGTCCGCCTGCCGC 1686
Qy 1381 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1440
Db 1687 AACATCGAGGCTTTCCGGATAGGATGGGGCGTTTTGCAATACGAGGAAAATGTCACCAAT 1746
Qy 1441 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1500
Db 1747 CCAGAAGATATGAGACCATATTGCTGGCACTACCCACCAAAACAGTGCGGCATAGTCCCC 1806
Qy 1501 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1560
Db 1807 GCGAGGTCTGTATGTGGCCCGGTGTACTGTTTCACTCCCAGCCCAGTGGTTGTGGGCACG 1866
Qy 1561 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1620
Db 1867 ACCGATAAGCTTGGAGTGCCTACTTATACATGGGGAGAGAATGAGACAGATGTCTTCCTA 1926
Qy 1621 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 1680
Db 1927 TTGAATAGCACCCGGCCACCGCGGGGGGCATGGTTCGGCTGCACGTGGATGAACTCCACT 1986
Qy 1681 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 1740
Db 1987 GGTTTCACCAAGACTTGTGGCGCACCACCTTGCCGCATCAGAGCTGACTTCAACGCCAGT 2046
Qy 1741 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 1800
Db 2047 ACGGACCTGTTATGCCCCACGGACTGTTTTAGGAAGCATCCTGAGGCCACTTACACCAAA 2106
Qy 1801 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 1860
Db 2107 TGTGGCTCTGGGCCTTGGCTCACGCCAAGATGCATGGTCGACTACCCCTACAGGCTCGGC 2166
Qy 1861 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 1920
Db 2167 ATTCATCCCTGCACAGTCAACTATACCATCTTCAAGATAAGGATGTATGTGGGGGGGGTT 2226
Qy 1921 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 1980
Db 2227 GAACACAGGCTTACGGCCGCGTGTAACTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 2286
Qy 1981 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2040
Db 2287 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTGCCT 2346
Qy 2041 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2100
Db 2347 TGTACTTTCTCAGACTTGCCCGCTTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2406
Qy 2101 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2160
Db 2407 GTGGACGTGCAATACATGTATGGCCTGTCACCTGCCATCACAAGACATATCGTTCGATGG 2466
Qy 2161 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2220
Db 2467 GAGTGGGTGGTACTCTTATTCCTGCTCTTGGCGGACGCCAGGGTCTGCGCATGCGTGTGG 2526
Qy 2221 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2280
Db 2527 ATGCTCATCTTGCTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCGTCTTGCACGCT 2586
Qy 2281 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATATTTTGTCATCTTTTTCGTGGCTGCTTGG 2340
Db 2587 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATACTTTGTCATCTTCTTCGTGGCTGCCTGG 2646
Qy 2341 TACATCAAGGGTCGGGTAGTCCCCTTAGCTACCTATTCCCTCACTGGCCTGTGGTCCTTT 2400
Db 2647 TACATCAAGGGGCGGGCAGTCCCCCTTGCTGCCTATTCCCTTACCGGCCTATGGTCCTTC 2706
Qy 2401 AGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTGCATGGC 2460
Db 2707 TGCCTACTGCTCCTAGCACTGCCCCAACAGGCTTATGCCTATGACGCATCTGTGCACGGG 2766
Qy 2461 CAGATAGGAGCGGCTCTGCTGGTAATGATCACTCTCTTTACTCTCACCCCCGGGTATAAG 2520
Db 2767 CAGATAGGCGCTGTTTTGTTGATACTAGTNACCCTTTTTACACTCACCCCGGGGTATAAG 2826
Qy 2521 ACCCTTCTCAGCCGGTTTTTGTGGTGGTTGTGCTATCTTCTGACCCTGGGGGAAGCTATG 2580
Db 2827 ATCCTTCTTAGCCGGTGCCTATGGTGGTTGTGCTATCTCCTGACCTTGGGGGAAGCCATG 2886
Qy 2581 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGCATCATATGGGCC 2640
Db 2887 ATCCAGGAGTGGGTACCACCCATGCAGGTGCGTGGCGGCCGCGATGGCATCATATGGGCC 2946
Qy 2641 GTCGCCATATTCTACCCAGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 2700
Db 2947 GCCACCATATTCTGCCCGGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGTTGCTT 3006
Qy 2701 GGGCCTGCTTACCTCCTAAAAGGTGCTTTGACGCGCGTGCCGTACTTCGTCAGGGCTCAC 2760
Db 3007 GGGCCTGGTTATCTCCTAAGAGGTGCTTTGACGCGTGTGCCGTATTTCGTCAGAGCTCAC 3066
Qy 2761 GCTCTACTGAGGATGTGCACCATGGCAAGGCATCTCGCGGGGGGCAGGTACGTCCAGATG 2820
Db 3067 GCTCTGCTGAGGATGTGCACCATGGCGAGGCACCTCGCGGGGGGTAAGTACGTCCAGATG 3126
Qy 2821 GCGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 2880
Db 3127 GCGTTATTAGCCCTTGGTAGGTGGACTGGCACTTACATCTATGACCATCTCACCCCTATG 3186
Qy 2881 TCGGATTGGGCTGCTAGTGGCCTGCGGGACCTGGCGGTCGCCGTTGAGCCTATCATCTTC 2940
Db 3187 TCGGACTGGGCTGCTAGTGGCCTGCGGGACCTGGCGGTCGCTGTGGAGCCCATCATCTTC 3246
Qy 2941 AGTCCGATGGAGAAGAAAGTCATTGTCTGGGGAGCGGAGACAGCTGCTTGTGGGGACATT 3000
Db 3247 AGTCCGATGGAGAAGAAAGTCATCGTCTGGGGAGCGGAGACAGCCGCATGCGGGGACATC 3306
Qy 3001 TTACACGGACTTCCCGTGTCCGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3060
Db 3307 CTACACGGACTCCCCGTGTCCGCCCGACTCGGCCGGGAGGTCCTCCTTGGCCCAGCTGAT 3366
Qy 3061 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3120
Db 3367 GGCTACACCTCCAAGGGGTGGAAGCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3426
Qy 3121 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCT 3180
Db 3427 CGAGGTCTCTTGGGTGCCATAGTGGTAAGCATGACGGGGCGAGACAAGACAGAACAGGCC 3486
Qy 3181 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3240
Db 3487 GGGGAGATCCAAGTCCTGTCCACAGTCACTCAATCCTTCCTCGGAACATCCATTTCAGGG 3546
Qy 3241 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3300
Db 3547 GTCTTATGGACTGTCTACCACGGAGCTGGCAACAAGACTCTAGCCGGCTCACGGGGTCCG 3606
Qy 3301 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGG 3360
Db 3607 GTCACACAGATGTACTCGAGTGCCGAGGGGGACTTGGTAGGGTGGCCCAGCCCTCCTGGG 3666
Qy 3361 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3420


Db 3667 ACCAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3726
Qy 3421 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3480
Db 3727 GCTGATGTCATCCCGGCTCGCAGGCGCGGGGACAAGCGGGGAGCGTTACTCTCCCCGAGA 3786
Qy 3481 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3540
Db 3787 CCCCTTTCGACCTTGAAGGGGTCCTCGGGGGGACCGGTGCTTTGCCCTAGGGGCCACGCT 3846
Qy 3541 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3600
Db 3847 GTCGGGATCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3906
Qy 3601 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 3660
Db 3907 CCCGTTGAGACACTTGACATCGTCACGCGGTCTCCCACCTTTAGTGACAACAGCACACCA 3966
Qy 3661 CCTGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 3720
Db 3967 CCGGCTGTGCCCCAGACCTATCAGGTCGGGTACTTACATGCCCCAACTGGCAGTGGAAAG 4026
Qy 3721 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 3780
Db 4027 AGCACCAAAGTCCCTGTCGCCTACGCCGCTCAGGGGTATAAGGTACTAATTCTTAATCCC 4086
Qy 3781 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 3840
Db 4087 TCGGTGGGTGCCACCTTGGGGTTTGGGGCGTACTTGTCCAAGGCATATGGCATCAATCCC 4146
Qy 3841 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 3900
Db 4147 AACATCAGGACTGGGGTCAGAACTGTGACAACCGGGGACTCCATCACATACTCCACATAT 4206
Qy 3901 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 3960
Db 4207 GGTAAATTCCTCGCTGATGGGGGCTGCGCAGGCGGCGCCTATGACATCATCATATGCGAT 4266
Qy 3961 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4020
Db 4267 GAATGCCACGCCGTTGGATCTACCACCAATCTCGGTATCGGAACAGTCCTTGATCAAGCA 4326
Qy 4021 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4080
Db 4327 GAGACGGCCGGGGTCAGACTAACTGTGCTGGCTACGGCTACGCTCCCCGGGTCAGTGGAG 4386
Qy 4081 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4140
Db 4387 ACCCCCCATCCCAATATAGAGGAGGTAGCCCTCGGGCAGGAGGGTGAGATCCCCTTCTAT 4446
Qy 4141 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4200
Db 4447 GGGAGGGCGATTCCCCTGTCCTACATCAAGGGAGGGAGGCACTTGATCTTCTGCCACTCA 4506
Qy 4201 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4260
Db 4507 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGCATGGGCTTGAACGCTGTGGCA 4566
Qy 4261 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4320
Db 4567 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGATGTAGTGGTCGTCGCC 4626
Qy 4321 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4380
Db 4627 ACCGACGCCCTCATGACGGGGTTCACTGGAGACTTTGACACCGTGATCGACTGCAACGTA 4686
Qy 4381 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4440
Db 4687 GCGGTCACTCAAGTCGTAGACTTCAGCCTGGACCCCACCTTCACCATAACCACACAGACT 4746
Qy 4441 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4500
Db 4747 GTACCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGGCGCACGGGTAGAGGAAGACTG 4806
Qy 4501 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4560
Db 4807 GGCACTTATAGGTACGTTTCCACTGGTGAACGAGCTTCAGGAATGTTTGACAGTGTAGTG 4866
Qy 4561 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4620
Db 4867 CTCTGTGAGTGCTACGACGCAGGAGCTGCATGGTATGAGCTTACACCAGCGGAGACCACC 4926
Qy 4621 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 4680
Db 4927 GTCAGGCTTAGAGCTTATTTTAACACACCTGGCTTGCCTGTGTGCCAAGACCATCTTGAG 4986
Qy 4681 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 4740
Db 4987 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 5046
Qy 4741 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 4800
Db 5047 AAGCAAGCGGGAGAGAGCTTTCCATACCTGGTAGCTTATCAAGCTACAGTATGCGCTAGA 5106
Qy 4801 GCCAAAGCCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 4860
Db 5107 GCCAAAGCCCCCCCCCCGTCCTGGGACATTATGTGGAAGTGCTTAACTCGACTCAAGCCC 5166
Qy 4861 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 4920
Db 5167 ACGCTCGTGGGCCCTACACCCCTCCTGTACCGTTTGGGCTCTGTTACCAACGAGATCACC 5226
Qy 4921 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 4980
Db 5227 CTCACACATCCTGTGACGAAATACATCGCCACATGCATGCAAGCTGACCTTGAGGTCATG 5286
Qy 4981 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5040
Db 5287 ACCAGCACGTGGGTCCTAGCTGGGGGAGTCTTGGCAGCCGTCGCCGCGTATTGCCTGGCG 5346
Qy 5041 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5100
Db 5347 ACCGGGTGTGTTTCCATCATCGGCCGCTTGCATATCAATCAGCGAGCCATCGTTGCACCG 5406
Qy 5101 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5160
Db 5407 GACAAGGAAGTCCTTTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAAAGCGGCT 5466
Qy 5161 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5220
Db 5467 CTCCTAGAAGAAGGGCAGCGAATAGCCGAGATGTTGAAGTCCAAGATCCAAGGCTTACTG 5526
Qy 5221 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5280
Db 5527 CAGCAAGCCTCTAAACAGGCCCAGGACATACAACCCGCTGTGCAGGCTTCATGGCCCAAG 5586
Qy 5281 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5340
Db 5587 GTGGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTAGCA 5646
Qy 5341 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5400
Db 5647 AGACTGTCAACACTGCCAGGGAACCCCGCTGTGGCTTCCATGATGGCGTTCAGTGCCGCC 5706
Qy 5401 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5460
Db 5707 CTCACCAGCCCGTTGTCAACTAGCACCACTATCCTTCTCAACATTTTGGGAAGGTGGCTG 5766
Qy 5461 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5520
Db 5767 GCGTCCCAAATTGCTCCACCCGCGGGGGCCACCGGCTTCGTCGTCAGTGGCCTGGTGGGG 5826
Qy 5521 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5580
Db 5827 GCTGCTGTGGGCAATATAGGTCTGGGTAAGGTGCTGGTGGACATTCTGGCAGGGTATGGC 5886


Qy 5581 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5640
Db 5887 GCGGGCATTTCGGGGGCCCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5946
Qy 5641 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 5700
Db 5947 GAAGATGTCATCAACCTGCTGCCTGGGATCCTGTCCCCGGGTGCCCTGGTGGTGGGGGTC 6006
Qy 5701 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 5760
Db 6007 ATCTGCGCGGCCATTCTGCGCCGCCACGTGGGACCGGGGGAAGGCGCGGTCCAATGGATG 6066
Qy 5761 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 5820
Db 6067 AACAGGCTCATCGCCTTCGCTTCCAGAGGAAACCACGTCGCCCCTACTCACTACGTGACG 6126
Qy 5821 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 5880
Db 6127 GAGTCAGATGCGTCGCAGCGTGTGACCCAGCTGCTTGGCTCTCTTACAATAACTAGCCTA 6186
Qy 5881 CTCAGAAGACTCCACAACTGGATTACTGAGGACTGCCCCATCCCATGCGGCGGCTCGTGG 5940
Db 6187 CTCAGGAGACTCCACAATTGGATTACTGAGGACTGCCCCATCCCATGCGCCGGCTCGTGG 6246
Qy 5941 CTCCGCGATGTGTGGGACTGGGTTTGCACCATCCTAACAGACTTTAAAAATTGGCTGACC 6000
Db 6247 CTTCGCGATGTGTGGGACTGGGTTTGCACCATCCTGACAGATTTTAAAAACTGGCTGACT 6306
Qy 6001 TCCAAATTATTCCCAAAGATGCCCGGCCTCCCCTTTGTCTCCTGTCAAAAGGGGTACAAG 6060
Db 6307 TCCAAGTTGTTCCCAAAGATGCCTGGCCTCCCCTTTATCTCTTGCCAAAAGGGGTACAAG 6366
Qy 6061 GGCGTGTGGGCCGGCACTGGCATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6120
Db 6367 GGTGTGTGGGCCGGCACTGGTATCATGACCACACGATGTCCCTGCGGCGCCAACATTTCT 6426
Qy 6121 GGCAATGTCCGCTTGGGCTCCATGAGAATCACGGGGCCTAAGACCTGCATGAATATCTGG 6180
Db 6427 GGCAATGTCCGCCTGGGCTCAATGAGAATCACAGGGCCTAAAACCTGCATGAACATTTGG 6486
Qy 6181 CAGGGGACCTTTCCTATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCGCCA 6240
Db 6487 CAGGGGACCTTTCCTATCAACTGTTACACTGAGGGCCAGTGCGTGCCGAAACCCGCGCCG 6546
Qy 6241 AACTTTAAGGTCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6300
Db 6547 AACTTCAAGACCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6606
Qy 6301 GGGTCATACCACTACATAACAGGACTCACCACTGATAACTTGAAAGTCCCCTGCCAACTA 6360
Db 6607 GGGCAATACTCCTATGTGACAGGATTGACCACTGACAACTTGAAAGTTCCCTGCCAACTA 6666
Qy 6361 CCCTCTCCCGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCACA 6420
Db 6667 CCTTCTCCAGAGTTCTTTTCCTGGGTGGACGGAGTGCAAATCCATAGGTTTGCCCCCACA 6726
Qy 6421 CCGAAGCCGTTTTTCCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6480
Db 6727 CCAAAGCCGTTCTTTCGGGATGAGGTCTCGTTTTATGTTGGGCTCAATTCGTTTGTCGTA 6786
Qy 6481 GGGTCCCAGCTTCCTTGCGACCCTGAACCCGACACAGACGTATTGATGTCCATGCTAACA 6540
Db 6787 GGGTCTCAGCTCCCCTGTGATCCTGAGCCCGATACAGATGTGTTGATGTCCATGTTAACA 6846
Qy 6541 GATCCATCTCATATCACGGCGGAGACTGCAGCGCGGCGTTTAGCGCGGGGGTCACCCCCA 6600
Db 6847 GAACCATCCCATATTACGGCGGAGGCTGCAGCGCGGCGCTTAGCACGGGGGTCACCCCCG 6906
Qy 6601 TCCGAGGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 6660
Db 6907 TCTGAGGCGAGCTCTTCAGCGAGCCAGCTATCAGCACCATCGCTGCGAGCCACCTGCACC 6966
Qy 6661 ACCCACGGCAAAGCCTATGATGTGGACATGGTGGATGCTAACCTGTTCATGGGGGGCGAT 6720
Db 6967 ACCCACGGCAAGGCTTATGATGTGGACATGGTGGATGCCAACCTGTTCATGGGGGGCGAT 7026
Qy 6721 GTGACTCGGATAGAGTCTGGGTCCAAAGTGGTCGTTCTGGACTCTCTCGACCCAATGGTC 6780
Db 7027 GTGACTCGGATAGAATCTGAGTCAAAAGTGGTCGTTCTGGACTCTCTCGACCCAATGGCT 7086
Qy 6781 GAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCAGAATACATGCTCCCCAAGAAGAGG 6840
Db 7087 GAAGAAAAGAGCGACCTCGAGCCCTCAATACCATCGGAGTATATGCTCCCCAGGAACAGG 7146
Qy 6841 TTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCGCTTGTGGAATCG 6900
Db 7147 TTCCCGCCAGCCTTACCGGCCTGGGCACGGCCTGATTACAACCCACCGCTTGTGGAATCG 7206
Qy 6901 TGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTCCCTCCTCCTAGG 6960
Db 7207 TGGAAGAGGCCAGATTACCAACCGCCCACTGTTGCGGGCTGTGCTCTCCCCCCCCCTAAG 7266
Qy 6961 AAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGTGAGGACTCCATA 7020
Db 7267 AAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGTCTGGATGAGAGCACCATA 7326
Qy 7021 GGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCCCCCAAGCGGCGAT 7080
Db 7327 GGAGATGCCCTCCAGCAGCTGGCCATCAAGACCTTCGGCCAGATCCCCCCAAGCGCCGAT 7386
Qy 7081 TCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACGCCTCCTGATGAG 7140
Db 7387 TCAGGCCTTTCCACGGGGGCGGACGCCGCCGACTCCGGCGGCCGGACGTCCCCTGACGAG 7446
Qy 7141 TTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCTCGAGGGGGAGCTTGGA 7200
Db 7447 TTGGCTCTTTCGGAGACAGGTTCCATTTCCTCCATGCCCCCCCTTGAGGGGGAGCCTGGG 7506
Qy 7201 GATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCCAGGGGGGGGTGGCA 7260
Db 7507 GATCCAGACCTAGAGCCCGAGCAGGTAGAGCTTCAACCTCCCCCCCAGGAGGGGGAGCTA 7566
Qy 7261 GCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGACGACTCCGTCGTG 7320
Db 7567 GCTCCCGGCTCGGACTCGGGGTCTTGGTCGACTTGCTCCGAGGAGGATGACTCCGTCGTG 7626
Qy 7321 TGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGTAGTCCCGAAGAG 7380
Db 7627 TGCTGCTCCATGTCGTACTCCTGGACCGGGGCTCTAATAACTCCTTGTAGCCCCGAAGAA 7686
Qy 7381 GAGAAGTTACCGATTAACCCCTTGAGCAACTCCCTGTTGCGATATCACAACAAGGTGTAC 7440
Db 7687 GAGAAGTTACCAATTAACCCCTTGAGCAACTCGTTGTTGCGATATCACAACAAGGTATAC 7746
Qy 7441 TGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTTGATAGGATGCAA 7500
Db 7747 TGTACTACATCAAAGAGTGCCTCACTGAGGGCTAAAAAGGTAACCTTTGACAGGATGCAA 7806
Qy 7501 GTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCGGCCTCCAAGGTC 7560
Db 7807 GTGCTCGACGCCCATTACGACTCAGTCTTAAAGGACATCAAGCTAGCGGCCTCTAAGGTC 7866
Qy 7561 ACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCCCATTCTGCAAGA 7620
Db 7867 AGTGCAAGGCTCCTCACCTTAGAGGAGGCCTGCCAGTTGACCCCACCCCATTCTGCAAGA 7926
Qy 7621 TCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGGGCCGTTAACCAC 7680
Db 7927 TCCAAATATGGATTTGGGGCTAAAGAGGTCCGCAGCTTGTCCGGGAGAGCCGTCAACCAC 7986
Qy 7681 ATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATTCCCACAACCATT 7740
Db 7987 ATCAAGTCCGTGTGGAAGGACCTCCTGGAAGACTCACAAACACCAATTCCTACGACCATC 8046
Qy 7741 ATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAGAAAGCAGCTCGC 7800

```
Db       8047 ATGGCCAAAAATGAGGTGTTCTGCGTCGACCCCACCAAGGGGGGTAAGAAAGCAGCTCGC 8106

Qy       7801 CTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCCCTTTATGACATT 7860

Db       8107 CTTATCGTTTACCCTGACCTCGGCGTCAGGGTATGCGAGAAAATGGCCCTTTATGATATC 8166

Qy       7861 ACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAGTATTCCCCCGCT 7920

Db       8167 ACACAAAAGCTTCCTCAGGCGGTGATGGGGGCTTCTTATGGCTTCCAGTACTCCCCTGCA 8226

Qy       7921 CAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCTATGGGTTTTTCG 7980

Db       8227 CAACGGGTGGAGTTTCTCTTGAAGGCATGGGCGGAAAAGAAAGACCCATTGGGTTTTTCG 8286

Qy       7981 TATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGGACTGAGGAGTCC 8040

Db       8287 TATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGAACTGAGGAGTCC 8346

Qy       8041 ATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACACTCGCTAACTGAG 8100

Db       8347 ATATACCGGGCCTGCTCCCTGCCCGAGGAAGCTCACACTGCCATACATTCGCTGACTGAG 8406

Qy       8101 AGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGCGGGTACAGGCGT 8160

Db       8407 AGACTTTACGTGGGGGGGCCTATGCTCAACAGCAAAGGCCAATCCTGCGGGTACAGGCGT 8466

Qy       8161 TGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACATGCTACGTGAAA 8220

Db       8467 TGCCGCGCCAGCGGAGTGCTCACCACTAGCATGGGGAACACCATCACATGCTATGTGAAA 8526

Qy       8221 GCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTGGTATGCGGCGAT 8280

Db       8527 GCCCTAGCAGCTTGTAAGGCTGCGGGGATAGTTGCGCCCACAATGCTGGTATGCGGCGAC 8586

Qy       8281 GACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGGAACCTGAGAGCC 8340

Db       8587 GACTTGGTTGTCATCTCAGAAAGCCAGGGGACTGAGGAGGACGAGCGGAACCTGAGAGCC 8646

Qy       8341 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCCAGACCGGAGTAT 8400

Db       8647 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCTCCCAGACCGGAATAT 8706

Qy       8401 GATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTGGGCCCACAAGGC 8460

Db       8707 GACCTGGAGCTGATAACATCTTGTTCCTCAAATGTGTCCGTGGCGCTGAACCCACTGGGC 8766

Qy       8461 CGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGGGCTGCCTGGGAA 8520

Db       8767 CGCCGCAGATACTACCTGACCAGAGACCCCACCACTCCAATTGCCCGGGCTGCCTGGGAA 8826

Qy       8521 ACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAGTACGCCCCGACC 8580

Db       8827 ACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAGTACGCCCCAACA 8886

Qy       8581 ATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATGGCTCAAGACACG 8640

Db       8887 ATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATGGCTCAAGACACC 8946

Qy       8641 CTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTGAGTCCCTTGGAC 8700

Db       8947 CTGGACCAGAATCTCAACTTTGAGATGTACGGATCGGTGTACTCCGTGAGTCCTTTGGAC 9006

Qy       8701 CTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTGCACACATACACT 8760

Db       9007 CTCCCAGCCATAATTGAAAGGCTACACGGGCTTGACGCTTTCTCTCTGCACACATACTCT 9066

Qy       8761 CCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGA 8820

Db       9067 CCCCACGAACTGACGCGGGTGGCTTCAACCCTCAGAAAACTTGGGGCGCCACCCCTCAGA 9126

Qy       8821 GCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGTGGGGGGAGAGCG 8880

Db       9127 GCGTGGAAGAGTCGGGCGCGTGCAGTTAGGGCGTCCCTCATCTCNCGTGGAGGGAGAGCG 9186

Qy       8881 GCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 8940

Db       9187 GCCGTTTGCGGTCGGTATCTCTTCAACTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 9246

Qy       8941 TTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGCGCCGGCGGGGGC 9000

Db       9247 TTGCCGGAGGCACGCCTCCTGGATTTATCCAGTTGGTTCACCGTCGGCGCCGGCGGGGGC 9306

Qy       9001 GACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTTGGCCTACTCCTA 9060

Db       9307 GACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTACTCCTTGGCCTACTCCTA 9366

Qy       9061 CTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCGG 9099

Db       9367 CTTTCTGTAGGGGTAGGCCTCTTTCTACTCCCCGCTCGG 9405
```

Search completed: October 31, 2005, 23:17:15
Job time : 25011 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 31, 2005, 14:42:35 ; Search time 2859 Seconds
(without alignments)
18840.058 Million cell updates/sec

Title: US-09-980-559-1_COPY_341_9439
Perfect score: 9099
Sequence: 1 atgagcacaaatcctaaacc.....ttttcctactccccgctcgg 9099

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	9099	100.0	9711	4	AAC86937	Aac86937 Nucleotid
2	9099	100.0	9711	5	AAC86644	Aac86644 Nucleotid
3	9095.8	100.0	9711	4	AAF23486	Aaf23486 Infectiou
4	8494.2	93.4	9589	2	AAQ38218	Aaq38218 NANBH vir
5	7513.4	82.6	9678	6	ABK88904	Abk88904 Human HCV
6	5652.2	62.1	9511	2	AAQ38221	Aaq38221 NANBH vir
7	5399.2	59.3	9611	5	AAC86645	Aac86645 Nucleotid
8	5399.2	59.3	9611	5	AAC86647	Aac86647 Nucleotid
9	5274.4	58.0	9611	5	AAC86646	Aac86646 Nucleotid
10	5274.4	58.0	9611	5	AAC86648	Aac86648 Nucleotid
11	4160.2	45.7	9436	2	AAQ63499	Aaq63499 Blood tra
12	4141.8	45.5	9413	6	AAD25517	Aad25517 Hepatitis
13	4141.8	45.5	9413	8	AAD49655	Aad49655 Hepatitis
14	4141.8	45.5	9413	10	ADF88596	Adf88596 Hepatitis
15	4140.2	45.5	9413	8	AAL53723	Aal53723 Hepatitis
16	4138.6	45.5	9413	2	AAQ81559	Aaq81559 Hepatitis
17	4138.6	45.5	9413	2	AAT03960	Aat03960 Partial H
18	4136	45.5	9599	2	AAX24833	Aax24833 Infectiou
19	4131.2	45.4	9611	13	ADS34713	Ads34713 Hepatitis
20	4131.2	45.4	12980	6	ABK87286	Abk87286 Hepatitis
21	4130.6	45.4	9413	2	AAQ80498	Aaq80498 DNA encod
22	4129.6	45.4	12980	2	AAV59364	Aav59364 Hepatitis
23	4129.6	45.4	12980	8	ACA62469	Aca62469 DNA encod
24	4128	45.4	9618	11	ADN33102	Adn33102 Hepatitis
25	4128	45.4	9646	2	AAV59361	Aav59361 Hepatitis
26	4128	45.4	9646	6	ABK87285	Abk87285 cDNA enco
27	4128	45.4	9646	8	ACA62466	Aca62466 HCV H77 c
28	4121.6	45.3	9599	2	AAX24832	Aax24832 Infectiou
29	4121.6	45.3	9599	4	AAC86938	Aac86938 Nucleotid
30	4121.6	45.3	9599	12	ADJ56743	Adj56743 Hepatitis
31	4121.6	45.3	9599	12	ADJ64255	Adj64255 Hepatitis
32	4121.6	45.3	10803	10	ABX10617	Abx10617 MKO-Z nuc
33	4120	45.3	9599	4	AAF23491	Aaf23491 Infectiou
34	4118.4	45.3	10803	10	ADD67945	Add67945 Modified
35	4115.2	45.2	9401	2	AAT12710	Aat12710 Hepatitis
36	4115.2	45.2	9401	2	AAT99981	Aat99981 HCV polyp
37	4115.2	45.2	9401	2	AAV09989	Aav09989 HCV polyp
38	4115.2	45.2	9401	6	AAD35043	Aad35043 Hepatitis
39	4115.2	45.2	9401	12	ADL23106	Adl23106 Hepatitis
40	4115.2	45.2	9401	13	ADR29358	Adr29358 Hepatitis
41	4114	45.2	9587	13	ADR82189	Adr82189 Hepatitis
42	4112	45.2	9365	6	AAD25518	Aad25518 Hepatitis
43	4112	45.2	9400	2	AAQ21744	Aaq21744 Compiled
44	4107.6	45.1	9416	2	AAX59394	Aax59394 Non-A, no
45	4106	45.1	9416	2	AAQ20268	Aaq20268 Non-A, no

ALIGNMENTS

RESULT 1
AAC86937
ID AAC86937 standard; DNA; 9711 BP.
XX
AC AAC86937;
XX
DT 02-APR-2001 (first entry)
XX
DE Nucleotide sequence of a hepatitis C virus (HCV) clone genotype 2a.
XX
KW Chimeric virus; bovine viral diarrhoea virus; BVDV; hepatitis C virus;
KW HCV; vaccine; viral inhibitor; antiviral; ss.
XX
OS Hepatitis C virus.
XX
FH Key Location/Qualifiers
FT CDS 341. .9442
FT /*tag= a
XX
PN WO200075352-A2.
XX
PD 14-DEC-2000.
XX
PF 02-JUN-2000; 2000WO-US015527.
XX
PR 04-JUN-1999; 99US-0137817P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Nam J, Bukh J, Emerson SU, Purcell RH;
XX
DR WPI; 2001-071081/08.
DR P-PSDB; AAB31168.
XX
PT New nucleic acid comprising a chimeric bovine viral diarrhea virus genome
PT in which the (non-)structural region has been replaced by hepatitis C
PT virus (HCV) genome useful for treating or preventing HCV signs and
PT symptoms.
XX
PS Disclosure; Page 82-85; 97pp; English.
XX
CC The specification describes a nucleic acid comprising a chimeric virus
CC genome, specifically bovine viral diarrhoea virus (BVDV) genome in which

CC the (non-)structural region has been replaced by the (non-)structural
CC region of a hepatitis C virus (HCV) genome. The nucleic acids comprising
CC the chimeric virus and the chimeric virus are useful for identifying cell
CC lines capable of supporting the replication of these chimeric viruses, in
CC screening for neutralizing antibodies to HCV of different genotypes, in
CC the production of HCV-BVDV virions, for the development of inactivated or
CC attenuated vaccines to prevent HCV-BVDV in a mammal, in studying the
CC molecular properties of HCV indirectly in vitro, and in identifying
CC inhibitors of viral enzyme activity which would be useful as antiviral
CC agents. Formulations or compositions comprising the chimeric virions may
CC be used to treat or prevent the signs and symptoms of HCV. The present
CC sequence represents a HCV clone, which is used to construct chimeric
CC nucleic acids of the invention
XX
SQ Sequence 9711 BP; 1987 A; 2853 C; 2668 G; 2203 T; 0 U; 0 Other;

Query Match 100.0%; Score 9099; DB 4; Length 9711;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 9099; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 400

Qy         61 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 460

Qy        121 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 520

Qy        181 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 580

Qy        241 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 640

Qy        301 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        641 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 700

Qy        361 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        701 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 760

Qy        421 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        761 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 820

Qy        481 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        821 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 880

Qy        541 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        881 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 940

Qy        601 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        941 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 1000

Qy        661 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1001 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 1060

Qy        721 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1061 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 1120

Qy        781 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1121 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 1180

Qy        841 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1181 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 1240

Qy        901 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1241 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 1300

Qy        961 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1301 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1360

Qy       1021 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1361 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1420

Qy       1081 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1421 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1480

Qy       1141 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1481 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1540

Qy       1201 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1541 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1600

Qy       1261 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1601 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1660

Qy       1321 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1661 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1720

Qy       1381 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1721 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1780

Qy       1441 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1781 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1840

Qy       1501 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1841 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1900

Qy       1561 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1901 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1960

Qy       1621 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1961 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 2020

Qy       1681 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2021 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 2080

Qy       1741 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2081 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 2140

Qy       1801 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2141 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 2200

Qy       1861 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       2201 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 2260

Qy       1921 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2261 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 2320

Qy       1981 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2321 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2380

Qy       2041 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2381 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2440

Qy       2101 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2441 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2500

Qy       2161 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2501 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2560

Qy       2221 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2561 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2620

Qy       2281 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATATTTTGTCATCTTTTTCGTGGCTGCTTGG 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2621 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATATTTTGTCATCTTTTTCGTGGCTGCTTGG 2680

Qy       2341 TACATCAAGGGTCGGGTAGTCCCCTTAGCTACCTATTCCCTCACTGGCCTGTGGTCCTTT 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2681 TACATCAAGGGTCGGGTAGTCCCCTTAGCTACCTATTCCCTCACTGGCCTGTGGTCCTTT 2740

Qy       2401 AGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTGCATGGC 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2741 AGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTGCATGGC 2800

Qy       2461 CAGATAGGAGCGGCTCTGCTGGTAATGATCACTCTCTTTACTCTCACCCCCGGGTATAAG 2520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2801 CAGATAGGAGCGGCTCTGCTGGTAATGATCACTCTCTTTACTCTCACCCCCGGGTATAAG 2860

Qy       2521 ACCCTTCTCAGCCGGTTTTTGTGGTGGTTGTGCTATCTTCTGACCCTGGGGGAAGCTATG 2580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2861 ACCCTTCTCAGCCGGTTTTTGTGGTGGTTGTGCTATCTTCTGACCCTGGGGGAAGCTATG 2920

Qy       2581 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGCATCATATGGGCC 2640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2921 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGCATCATATGGGCC 2980

Qy       2641 GTCGCCATATTCTACCCAGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 2700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2981 GTCGCCATATTCTACCCAGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 3040

Qy       2701 GGGCCTGCTTACCTCCTAAAAGGTGCTTTGACGCGCGTGCCGTACTTCGTCAGGGCTCAC 2760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3041 GGGCCTGCTTACCTCCTAAAAGGTGCTTTGACGCGCGTGCCGTACTTCGTCAGGGCTCAC 3100

Qy       2761 GCTCTACTGAGGATGTGCACCATGGCAAGGCATCTCGCGGGGGGCAGGTACGTCCAGATG 2820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3101 GCTCTACTGAGGATGTGCACCATGGCAAGGCATCTCGCGGGGGGCAGGTACGTCCAGATG 3160

Qy       2821 GCGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 2880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3161 GCGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 3220

Qy       2881 TCGGATTGGGCTGCTAGTGGCCTGCGGGACCTGGCGGTCGCCGTTGAGCCTATCATCTTC 2940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3221 TCGGATTGGGCTGCTAGTGGCCTGCGGGACCTGGCGGTCGCCGTTGAGCCTATCATCTTC 3280

Qy       2941 AGTCCGATGGAGAAGAAAGTCATTGTCTGGGGAGCGGAGACAGCTGCTTGTGGGGACATT 3000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3281 AGTCCGATGGAGAAGAAAGTCATTGTCTGGGGAGCGGAGACAGCTGCTTGTGGGGACATT 3340

Qy       3001 TTACACGGACTTCCCGTGTCCGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3341 TTACACGGACTTCCCGTGTCCGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3400

Qy       3061 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3401 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3460

Qy       3121 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCT 3180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3461 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCT 3520

Qy       3181 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3521 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3580

Qy       3241 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3581 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3640

Qy       3301 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGG 3360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3641 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGG 3700

Qy       3361 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3701 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3760

Qy       3421 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3761 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3820

Qy       3481 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3821 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3880

Qy       3541 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3881 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3940

Qy       3601 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 3660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3941 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 4000

Qy       3661 CCTGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 3720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4001 CCTGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 4060

Qy       3721 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 3780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4061 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 4120

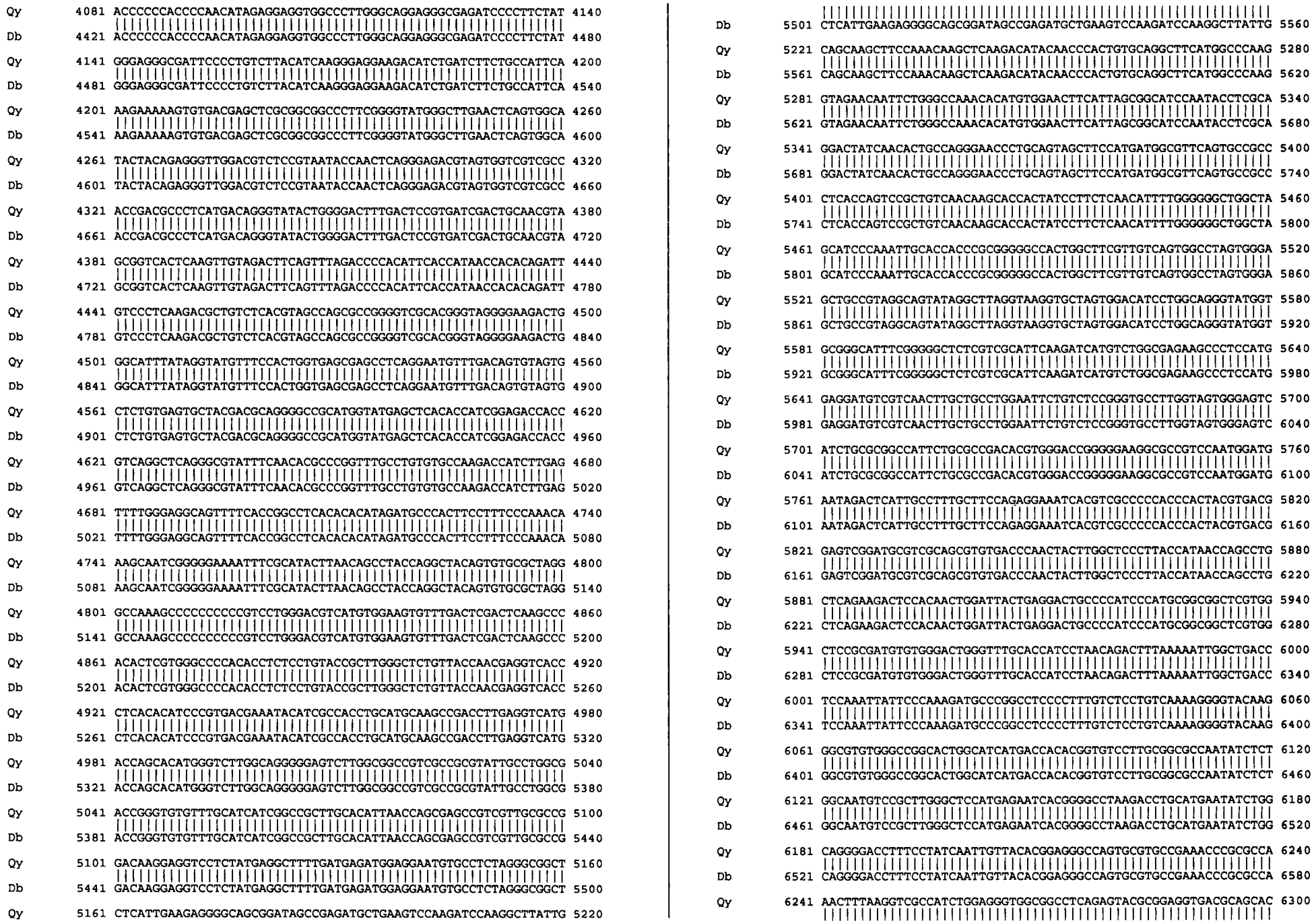
Qy       3781 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 3840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4121 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 4180

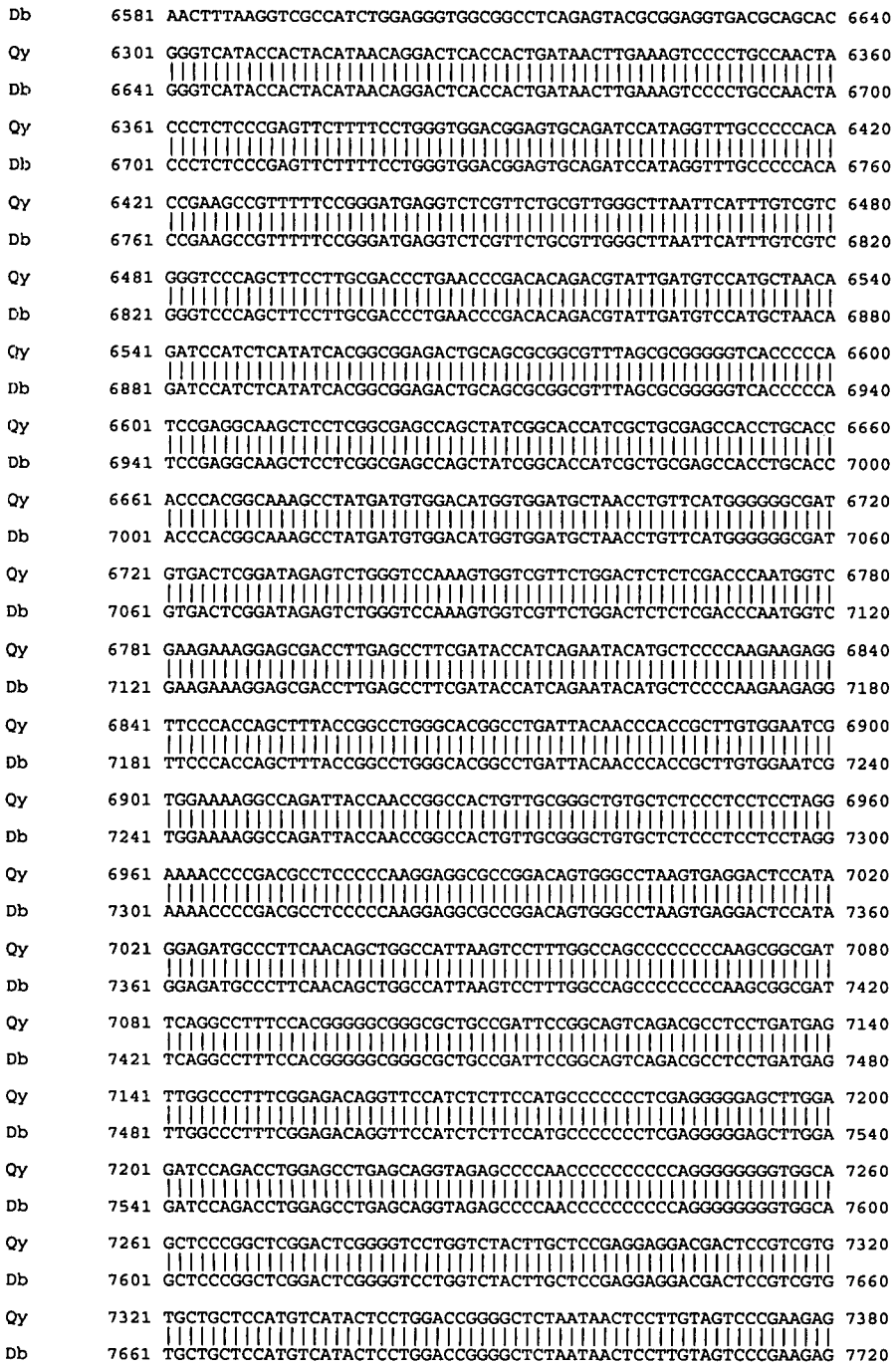
Qy       3841 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 3900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4181 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 4240

Qy       3901 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 3960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4241 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 4300

Qy       3961 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4301 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4360

Qy       4021 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4361 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4420
```


Qy 4081 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4140
Db 4421 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4480
Qy 4141 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4200
Db 4481 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4540
Qy 4201 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4260
Db 4541 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4600
Qy 4261 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4320
Db 4601 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4660
Qy 4321 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4380
Db 4661 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4720
Qy 4381 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4440
Db 4721 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4780
Qy 4441 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4500
Db 4781 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4840
Qy 4501 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4560
Db 4841 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4900
Qy 4561 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4620
Db 4901 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4960
Qy 4621 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 4680
Db 4961 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 5020
Qy 4681 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 4740
Db 5021 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 5080
Qy 4741 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 4800
Db 5081 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 5140
Qy 4801 GCCAAAGCCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 4860
Db 5141 GCCAAAGCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 5200
Qy 4861 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 4920
Db 5201 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 5260
Qy 4921 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 4980
Db 5261 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 5320
Qy 4981 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5040
Db 5321 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5380
Qy 5041 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5100
Db 5381 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5440
Qy 5101 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5160
Db 5441 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5500
Qy 5161 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5220
Db 5501 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5560
Qy 5221 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5280
Db 5561 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5620
Qy 5281 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5340
Db 5621 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5680
Qy 5341 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5400
Db 5681 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5740
Qy 5401 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5460
Db 5741 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5800
Qy 5461 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5520
Db 5801 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5860
Qy 5521 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5580
Db 5861 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5920
Qy 5581 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5640
Db 5921 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5980
Qy 5641 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 5700
Db 5981 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 6040
Qy 5701 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 5760
Db 6041 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 6100
Qy 5761 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 5820
Db 6101 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 6160
Qy 5821 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 5880
Db 6161 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 6220
Qy 5881 CTCAGAAGACTCCACAACTGGATTACTGAGGACTGCCCCATCCCATGCGGCGGCTCGTGG 5940
Db 6221 CTCAGAAGACTCCACAACTGGATTACTGAGGACTGCCCCATCCCATGCGGCGGCTCGTGG 6280
Qy 5941 CTCCGCGATGTGTGGGACTGGGTTTGCACCATCCTAACAGACTTTAAAAATTGGCTGACC 6000
Db 6281 CTCCGCGATGTGTGGGACTGGGTTTGCACCATCCTAACAGACTTTAAAAATTGGCTGACC 6340
Qy 6001 TCCAAATTATTCCCAAAGATGCCCGGCCTCCCCTTTGTCTCCTGTCAAAAGGGGTACAAG 6060
Db 6341 TCCAAATTATTCCCAAAGATGCCCGGCCTCCCCTTTGTCTCCTGTCAAAAGGGGTACAAG 6400
Qy 6061 GGCGTGTGGGCCGGCACTGGCATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6120
Db 6401 GGCGTGTGGGCCGGCACTGGCATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6460
Qy 6121 GGCAATGTCCGCTTGGGCTCCATGAGAATCACGGGGCCTAAGACCTGCATGAATATCTGG 6180
Db 6461 GGCAATGTCCGCTTGGGCTCCATGAGAATCACGGGGCCTAAGACCTGCATGAATATCTGG 6520
Qy 6181 CAGGGGACCTTTCCTATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCGCCA 6240
Db 6521 CAGGGGACCTTTCCTATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCGCCA 6580
Qy 6241 AACTTTAAGGTCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6300


Db 6581 AACTTTAAGGTCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6640
Qy 6301 GGGTCATACCACTACATAACAGGACTCACCACTGATAACTTGAAAGTCCCCTGCCAACTA 6360
Db 6641 GGGTCATACCACTACATAACAGGACTCACCACTGATAACTTGAAAGTCCCCTGCCAACTA 6700
Qy 6361 CCCTCTCCCGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCACA 6420
Db 6701 CCCTCTCCCGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCACA 6760
Qy 6421 CCGAAGCCGTTTTTCCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6480
Db 6761 CCGAAGCCGTTTTTCCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6820
Qy 6481 GGGTCCCAGCTTCCTTGCGACCCTGAACCCGACACAGACGTATTGATGTCCATGCTAACA 6540
Db 6821 GGGTCCCAGCTTCCTTGCGACCCTGAACCCGACACAGACGTATTGATGTCCATGCTAACA 6880
Qy 6541 GATCCATCTCATATCACGGCGGAGACTGCAGCGCGGCGTTTAGCGCGGGGGTCACCCCCA 6600
Db 6881 GATCCATCTCATATCACGGCGGAGACTGCAGCGCGGCGTTTAGCGCGGGGGTCACCCCCA 6940
Qy 6601 TCCGAGGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 6660
Db 6941 TCCGAGGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 7000
Qy 6661 ACCCACGGCAAAGCCTATGATGTGGACATGGTGGATGCTAACCTGTTCATGGGGGGCGAT 6720
Db 7001 ACCCACGGCAAAGCCTATGATGTGGACATGGTGGATGCTAACCTGTTCATGGGGGGCGAT 7060
Qy 6721 GTGACTCGGATAGAGTCTGGGTCCAAAGTGGTCGTTCTGGACTCTCTCGACCCAATGGTC 6780
Db 7061 GTGACTCGGATAGAGTCTGGGTCCAAAGTGGTCGTTCTGGACTCTCTCGACCCAATGGTC 7120
Qy 6781 GAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCAGAATACATGCTCCCCAAGAAGAGG 6840
Db 7121 GAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCAGAATACATGCTCCCCAAGAAGAGG 7180
Qy 6841 TTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCGCTTGTGGAATCG 6900
Db 7181 TTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCGCTTGTGGAATCG 7240
Qy 6901 TGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTCCCTCCTCCTAGG 6960
Db 7241 TGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTCCCTCCTCCTAGG 7300
Qy 6961 AAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGTGAGGACTCCATA 7020
Db 7301 AAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGTGAGGACTCCATA 7360
Qy 7021 GGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCCCCCAAGCGGCGAT 7080
Db 7361 GGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCCCCAAGCGGCGAT 7420
Qy 7081 TCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACGCCTCCTGATGAG 7140
Db 7421 TCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACGCCTCCTGATGAG 7480
Qy 7141 TTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCTCGAGGGGGAGCTTGGA 7200
Db 7481 TTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCTCGAGGGGGAGCTTGGA 7540
Qy 7201 GATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCCAGGGGGGGGTGGCA 7260
Db 7541 GATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCCAGGGGGGGGTGGCA 7600
Qy 7261 GCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGACGACTCCGTCGTG 7320
Db 7601 GCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGACGACTCCGTCGTG 7660
Qy 7321 TGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGTAGTCCCGAAGAG 7380
Db 7661 TGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGTAGTCCCGAAGAG 7720
Qy 7381 GAGAAGTTACCGATTAACCCCTTGAGCAACTCCCTGTTGCGATATCACAACAAGGTGTAC 7440
Db 7721 GAGAAGTTACCGATTAACCCCTTGAGCAACTCCCTGTTGCGATATCACAACAAGGTGTAC 7780
Qy 7441 TGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTTGATAGGATGCAA 7500
Db 7781 TGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTTGATAGGATGCAA 7840
Qy 7501 GTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCGGCCTCCAAGGTC 7560
Db 7841 GTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCGGCCTCCAAGGTC 7900
Qy 7561 ACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCCCATTCTGCAAGA 7620
Db 7901 ACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCCCATTCTGCAAGA 7960
Qy 7621 TCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGGGCCGTTAACCAC 7680
Db 7961 TCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGGGCCGTTAACCAC 8020
Qy 7681 ATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATTCCCACAACCATT 7740
Db 8021 ATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATTCCCACAACCATT 8080
Qy 7741 ATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAGAAAGCAGCTCGC 7800
Db 8081 ATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAGAAAGCAGCTCGC 8140
Qy 7801 CTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCCCTTTATGACATT 7860
Db 8141 CTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCCCTTTATGACATT 8200
Qy 7861 ACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAGTATTCCCCCGCT 7920
Db 8201 ACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAGTATTCCCCCGCT 8260
Qy 7921 CAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCTATGGGTTTTTCG 7980
Db 8261 CAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCTATGGGTTTTTCG 8320
Qy 7981 TATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGGACTGAGGAGTCC 8040
Db 8321 TATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGGACTGAGGAGTCC 8380
Qy 8041 ATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACACTCGCTAACTGAG 8100
Db 8381 ATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACACTCGCTAACTGAG 8440
Qy 8101 AGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGCGGGTACAGGCGT 8160
Db 8441 AGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGCGGGTACAGGCGT 8500
Qy 8161 TGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACATGCTACGTGAAA 8220
Db 8501 TGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACATGCTACGTGAAA 8560
Qy 8221 GCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTGGTATGCGGCGAT 8280
Db 8561 GCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTGGTATGCGGCGAT 8620
Qy 8281 GACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGGAACCTGAGAGCC 8340
Db 8621 GACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGGAACCTGAGAGCC 8680
Qy 8341 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCCAGACCGGAGTAT 8400
Db 8681 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCCAGACCGGAGTAT 8740
Qy 8401 GATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTGGGCCCACAAGGC 8460
Db 8741 GATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTGGGCCCACAAGGC 8800

```
Qy       8461 CGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGGGCTGCCTGGGAA 8520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8801 CGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGGGCTGCCTGGGAA 8860

Qy       8521 ACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAGTACGCCCCGACC 8580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8861 ACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAGTACGCCCCGACC 8920

Qy       8581 ATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATGGCTCAAGACACG 8640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8921 ATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATGGCTCAAGACACG 8980

Qy       8641 CTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTGAGTCCCTTGGAC 8700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8981 CTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTGAGTCCCTTGGAC 9040

Qy       8701 CTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTGCACACATACACT 8760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9041 CTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTGCACACATACACT 9100

Qy       8761 CCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGA 8820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9101 CCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGA 9160

Qy       8821 GCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGTGGGGGGAGAGCG 8880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9161 GCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGTGGGGGGAGAGCG 9220

Qy       8881 GCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 8940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9221 GCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 9280

Qy       8941 TTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGCGCCGGCGGGGGC 9000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9281 TTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGCGCCGGCGGGGGC 9340

Qy       9001 GACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTTGGCCTACTCCTA 9060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9341 GACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTTGGCCTACTCCTA 9400

Qy       9061 CTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCGG 9099
             |||||||||||||||||||||||||||||||||||||||
Db       9401 CTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCGG 9439
```

RESULT 2
AAC86644

```
ID   AAC86644 standard; DNA; 9711 BP.
XX
AC   AAC86644;
XX
DT   02-APR-2001  (first entry)
XX
DE   Nucleotide sequence of infectious Hepatitis C virus strain HC-J6CH.
XX
KW   HCV; HCV strain HC-J6CH; HCV genotype 2a; antiviral; vaccine; ss.
XX
OS   Hepatitis C virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             341. .9442
FT                   /*tag=  a
FT                   /product= "Hepatitis C virus protein"
XX
PN   WO200075338-A2.
XX
PD   14-DEC-2000.
XX
PF   02-JUN-2000; 2000WO-US015446.
XX
PR   04-JUN-1999;   99US-0137693P.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Yanagi M,  Bukh J,  Emerson SU,  Purcell RH;
XX
DR   WPI; 2001-061728/07.
DR   P-PSDB; AAB30729.
XX
PT   Nucleic acid molecule encoding human hepatitis C virus of genotype 2a for
PT   developing vaccines, for diagnosis of hepatitis C virus and in screening
PT   assays for identification of antiviral agents.
XX
PS   Claim 3; Page 84-87; 167pp; English.
XX
CC   The present sequence represents a nucleotide sequence of infectious
CC   Hepatitis C virus (HCV) strain HC-J6CH genotype 2a. The HCV
CC   polynucleotide sequence is capable of expressing the virus when
CC   transfected into cells. The HCV protein is useful for assaying candidate
CC   antiviral agents for activity against HCV. Antibodies specific for HCV
CC   polypeptide are useful in prevention and treatment of diseases caused by
CC   HCV in animals, in particular humans. The HCV polypeptides serve as
CC   immunogens in the development of vaccines for preventing HCV in mammals
CC   or as antigens in diagnostic assays for detecting the presence of HCV in
CC   biological samples. The HCV polynucleotide is also useful for identifying
CC   cell lines capable of supporting the replication of HCV in vitro and to
CC   produce attenuated viral strains via passage in vitro or in vivo
XX
SQ   Sequence 9711 BP; 1987 A; 2853 C; 2668 G; 2203 T; 0 U; 0 Other;
```

Query Match 100.0%; Score 9099; DB 5; Length 9711;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 9099; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 400

Qy         61 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 460

Qy        121 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 520

Qy        181 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 580

Qy        241 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 640

Qy        301 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        641 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 700

Qy        361 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        701 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 760

Qy        421 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        761 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 820

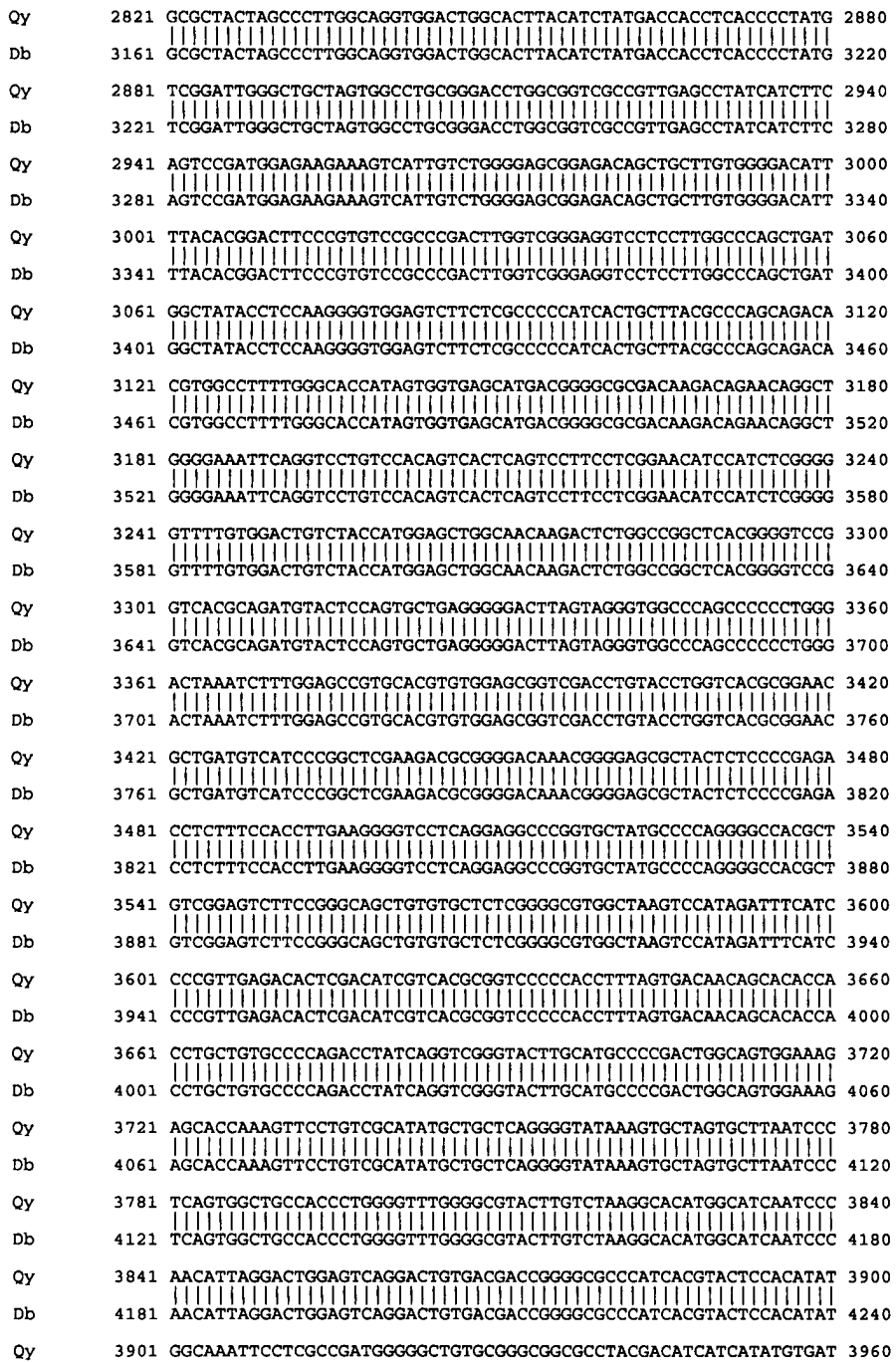
Qy        481 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        821 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 880

Qy        541 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        881 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 940

Qy        601 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        941 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 1000
Qy        661 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 720
Db       1001 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 1060
Qy        721 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 780
Db       1061 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 1120
Qy        781 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 840
Db       1121 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 1180
Qy        841 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 900
Db       1181 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 1240
Qy        901 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 960
Db       1241 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 1300
Cy        961 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1020
Db       1301 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1360
Qy       1021 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1080
Db       1361 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1420
Qy       1081 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1140
Db       1421 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1480
Qy       1141 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1200
Db       1481 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1540
Qy       1201 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1260
Db       1541 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1600
Qy       1261 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1320
Db       1601 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1660
Qy       1321 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1380
Db       1661 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1720
Qy       1381 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1440
Db       1721 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1780
Qy       1441 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1500
Db       1781 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1840
Qy       1501 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1560
Db       1841 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1900
Qy       1561 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1620
Db       1901 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1960
Qy       1621 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 1680
Db       1961 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 2020
Qy       1681 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 1740
Db       2021 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 2080

Qy       1741 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 1800
Db       2081 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 2140
Qy       1801 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 1860
Db       2141 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 2200
Qy       1861 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 1920
Db       2201 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 2260
Qy       1921 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 1980
Db       2261 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 2320
Qy       1981 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2040
Db       2321 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2380
Qy       2041 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2100
Db       2381 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2440
Qy       2101 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2160
Db       2441 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2500
Qy       2161 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2220
Db       2501 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2560
Qy       2221 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2280
Db       2561 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2620
Qy       2281 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATATTTTGTCATCTTTTTCGTGGCTGCTTGG 2340
Db       2621 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATATTTTGTCATCTTTTTCGTGGCTGCTTGG 2680
Qy       2341 TACATCAAGGGTCGGGTAGTCCCCTTAGCTACCTATTCCCTCACTGGCCTGTGGTCCTTT 2400
Db       2681 TACATCAAGGGTCGGGTAGTCCCCTTAGCTACCTATTCCCTCACTGGCCTGTGGTCCTTT 2740
Qy       2401 AGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTGCATGGC 2460
Db       2741 AGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTGCATGGC 2800
Qy       2461 CAGATAGGAGCGGCTCTGCTGGTAATGATCACTCTCTTTACTCTCACCCCCGGGTATAAG 2520
Db       2801 CAGATAGGAGCGGCTCTGCTGGTAATGATCACTCTCTTTACTCTCACCCCCGGGTATAAG 2860
Qy       2521 ACCCTTCTCAGCCGGTTTTTGTGGTGGTTGTGCTATCTTCTGACCCTGGGGGAAGCTATG 2580
Db       2861 ACCCTTCTCAGCCGGTTTTTGTGGTGGTTGTGCTATCTTCTGACCCTGGGGGAAGCTATG 2920
Qy       2581 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGCATCATATGGGCC 2640
Db       2921 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGCATCATATGGGCC 2980
Qy       2641 GTCGCCATATTCTACCCAGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 2700
Db       2981 GTCGCCATATTCTACCCAGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 3040
Qy       2701 GGGCCTGCTTACCTCCTAAAAGGTGCTTTGACGCGCGTGCCGTACTTCGTCAGGGCTCAC 2760
Db       3041 GGGCCTGCTTACCTCCTAAAAGGTGCTTTGACGCGCGTGCCGTACTTCGTCAGGGCTCAC 3100
Qy       2761 GCTCTACTGAGGATGTGCACCATGGCAAGGCATCTCGCGGGGGGCAGGTACGTCCAGATG 2820
Db       3101 GCTCTACTGAGGATGTGCACCATGGCAAGGCATCTCGCGGGGGGCAGGTACGTCCAGATG 3160
```


Qy 2821 GCGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 2880
Db 3161 GCGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 3220
Qy 2881 TCGGATTGGGCTGCTAGTGGCCTGCGGGACCTGGCGGTCGCCGTTGAGCCTATCATCTTC 2940
Db 3221 TCGGATTGGGCTGCTAGTGGCCTGCGGGACCTGGCGGTCGCCGTTGAGCCTATCATCTTC 3280
Qy 2941 AGTCCGATGGAGAAGAAAGTCATTGTCTGGGGAGCGGAGACAGCTGCTTGTGGGGACATT 3000
Db 3281 AGTCCGATGGAGAAGAAAGTCATTGTCTGGGGAGCGGAGACAGCTGCTTGTGGGGACATT 3340
Qy 3001 TTACACGGACTTCCCGTGTCCGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3060
Db 3341 TTACACGGACTTCCCGTGTCCGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3400
Qy 3061 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3120
Db 3401 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3460
Qy 3121 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCT 3180
Db 3461 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCT 3520
Qy 3181 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3240
Db 3521 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3580
Qy 3241 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3300
Db 3581 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3640
Qy 3301 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGG 3360
Db 3641 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGG 3700
Qy 3361 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3420
Db 3701 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3760
Qy 3421 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3480
Db 3761 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3820
Qy 3481 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3540
Db 3821 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3880
Qy 3541 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3600
Db 3881 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3940
Qy 3601 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 3660
Db 3941 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 4000
Qy 3661 CCTGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 3720
Db 4001 CCTGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 4060
Qy 3721 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 3780
Db 4061 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 4120
Qy 3781 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 3840
Db 4121 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 4180
Qy 3841 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 3900
Db 4181 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 4240
Qy 3901 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 3960
Db 4241 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 4300
Qy 3961 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4020
Db 4301 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4360
Qy 4021 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4080
Db 4361 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4420
Qy 4081 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4140
Db 4421 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4480
Qy 4141 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4200
Db 4481 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4540
Qy 4201 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4260
Db 4541 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4600
Qy 4261 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4320
Db 4601 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4660
Qy 4321 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4380
Db 4661 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4720
Qy 4381 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4440
Db 4721 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4780
Qy 4441 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4500
Db 4781 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4840
Qy 4501 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4560
Db 4841 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4900
Qy 4561 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4620
Db 4901 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4960
Qy 4621 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 4680
Db 4961 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 5020
Qy 4681 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 4740
Db 5021 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 5080
Qy 4741 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 4800
Db 5081 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 5140
Qy 4801 GCCAAAGCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 4860
Db 5141 GCCAAAGCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 5200
Qy 4861 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 4920
Db 5201 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 5260
Qy 4921 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 4980
Db 5261 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 5320
Qy 4981 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5040

```
Db       5321 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5380

Qy       5041 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5381 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5440

Qy       5101 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5441 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5500

Qy       5161 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5501 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5560

Qy       5221 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5561 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5620

Qy       5281 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5621 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5680

Qy       5341 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5681 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5740

Qy       5401 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5741 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5800

Qy       5461 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5801 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5860

Qy       5521 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5861 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5920

Qy       5581 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5921 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5980

Qy       5641 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 5700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5981 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 6040

Qy       5701 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 5760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6041 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 6100

Qy       5761 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 5820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6101 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 6160

Qy       5821 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 5880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6161 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 6220

Qy       5881 CTCAGAAGACTCCACAACTGGATTACTGAGGACTGCCCCATCCCATGCGGCGGCTCGTGG 5940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6221 CTCAGAAGACTCCACAACTGGATTACTGAGGACTGCCCCATCCCATGCGGCGGCTCGTGG 6280

Qy       5941 CTCCGCGATGTGTGGGACTGGGTTTGCACCATCCTAACAGACTTTAAAAATTGGCTGACC 6000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6281 CTCCGCGATGTGTGGGACTGGGTTTGCACCATCCTAACAGACTTTAAAAATTGGCTGACC 6340

Qy       6001 TCCAAATTATTCCCAAAGATGCCCGGCCTCCCCTTTGTCTCCTGTCAAAAGGGGTACAAG 6060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6341 TCCAAATTATTCCCAAAGATGCCCGGCCTCCCCTTTGTCTCCTGTCAAAAGGGGTACAAG 6400

Qy       6061 GGCGTGTGGGCCGGCACTGGCATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6401 GGCGTGTGGGCCGGCACTGGCATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6460

Qy       6121 GGCAATGTCCGCTTGGGCTCCATGAGAATCACGGGGCCTAAGACCTGCATGAATATCTGG 6180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6461 GGCAATGTCCGCTTGGGCTCCATGAGAATCACGGGGCCTAAGACCTGCATGAATATCTGG 6520

Qy       6181 CAGGGGACCTTTCCTATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCGCCA 6240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6521 CAGGGGACCTTTCCTATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCGCCA 6580

Qy       6241 AACTTTAAGGTCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6581 AACTTTAAGGTCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6640

Qy       6301 GGGTCATACCACTACATAACAGGACTCACCACTGATAACTTGAAAGTCCCCTGCCAACTA 6360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6641 GGGTCATACCACTACATAACAGGACTCACCACTGATAACTTGAAAGTCCCCTGCCAACTA 6700

Qy       6361 CCCTCTCCCGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCACA 6420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6701 CCCTCTCCCGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCACA 6760

Qy       6421 CCGAAGCCGTTTTTCCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6761 CCGAAGCCGTTTTTCCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6820

Qy       6481 GGGTCCCAGCTTCCTTGCGACCCTGAACCCGACACAGACGTATTGATGTCCATGCTAACA 6540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6821 GGGTCCCAGCTTCCTTGCGACCCTGAACCCGACACAGACGTATTGATGTCCATGCTAACA 6880

Qy       6541 GATCCATCTCATATCACGGCGGAGACTGCAGCGCGGCGTTTAGCGCGGGGGTCACCCCCA 6600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6881 GATCCATCTCATATCACGGCGGAGACTGCAGCGCGGCGTTTAGCGCGGGGGTCACCCCCA 6940

Qy       6601 TCCGAGGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 6660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6941 TCCGAGGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 7000

Qy       6661 ACCCACGGCAAAGCCTATGATGTGGACATGGTGGATGCTAACCTGTTCATGGGGGGCGAT 6720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7001 ACCCACGGCAAAGCCTATGATGTGGACATGGTGGATGCTAACCTGTTCATGGGGGGCGAT 7060

Qy       6721 GTGACTCGGATAGAGTCTGGGTCCAAAGTGGTCGTTCTGGACTCTCTCGACCCAATGGTC 6780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7061 GTGACTCGGATAGAGTCTGGGTCCAAAGTGGTCGTTCTGGACTCTCTCGACCCAATGGTC 7120

Qy       6781 GAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCAGAATACATGCTCCCCAAGAAGAGG 6840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7121 GAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCAGAATACATGCTCCCCAAGAAGAGG 7180

Qy       6841 TTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCGCTTGTGGAATCG 6900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7181 TTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCGCTTGTGGAATCG 7240

Qy       6901 TGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTCCCTCCTCCTAGG 6960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7241 TGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTCCCTCCTCCTAGG 7300

Qy       6961 AAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGTGAGGACTCCATA 7020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7301 AAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGTGAGGACTCCATA 7360

Qy       7021 GGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCCCCCAAGCGGCGAT 7080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7361 GGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCCCCCAAGCGGCGAT 7420

Qy       7081 TCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACGCCTCCTGATGAG 7140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7421 TCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACGCCTCCTGATGAG 7480

Qy       7141 TTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCTCGAGGGGGAGCTTGGA 7200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7481 TTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCTCGAGGGGGAGCTTGGA 7540
```

```
Qy       7201 GATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCCAGGGGGGGGTGGCA 7260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7541 GATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCCAGGGGGGGGTGGCA 7600

Qy       7261 GCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGACGACTCCGTCGTG 7320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7601 GCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGACGACTCCGTCGTG 7660

Qy       7321 TGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGTAGTCCCGAAGAG 7380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7661 TGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGTAGTCCCGAAGAG 7720

Qy       7381 GAGAAGTTACCGATTAACCCCTTGAGCAACTCCCTGTTGCGATATCACAACAAGGTGTAC 7440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7721 GAGAAGTTACCGATTAACCCCTTGAGCAACTCCCTGTTGCGATATCACAACAAGGTGTAC 7780

Qy       7441 TGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTTGATAGGATGCAA 7500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7781 TGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTTGATAGGATGCAA 7840

Qy       7501 GTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCGGCCTCCAAGGTC 7560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7841 GTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCGGCCTCCAAGGTC 7900

Qy       7561 ACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCCCATTCTGCAAGA 7620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7901 ACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCCCATTCTGCAAGA 7960

Qy       7621 TCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGGGCCGTTAACCAC 7680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7961 TCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGGGCCGTTAACCAC 8020

Qy       7681 ATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATTCCCACAACCATT 7740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8021 ATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATTCCCACAACCATT 8080

Qy       7741 ATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAGAAAGCAGCTCGC 7800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8081 ATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAGAAAGCAGCTCGC 8140

Qy       7801 CTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCCCTTTATGACATT 7860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8141 CTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCCCTTTATGACATT 8200

Qy       7861 ACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAGTATTCCCCCGCT 7920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8201 ACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAGTATTCCCCCGCT 8260

Qy       7921 CAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCTATGGGTTTTTCG 7980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8261 CAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCTATGGGTTTTTCG 8320

Qy       7981 TATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGGACTGAGGAGTCC 8040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8321 TATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGGACTGAGGAGTCC 8380

Qy       8041 ATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACACTCGCTAACTGAG 8100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8381 ATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACACTCGCTAACTGAG 8440

Qy       8101 AGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGCGGGTACAGGCGT 8160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8441 AGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGCGGGTACAGGCGT 8500

Qy       8161 TGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACATGCTACGTGAAA 8220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8501 TGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACATGCTACGTGAAA 8560

Qy       8221 GCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTGGTATGCGGCGAT 8280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8561 GCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTGGTATGCGGCGAT 8620

Qy       8281 GACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGGAACCTGAGAGCC 8340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8621 GACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGGAACCTGAGAGCC 8680

Qy       8341 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCCAGACCGGAGTAT 8400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8681 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCCAGACCGGAGTAT 8740

Qy       8401 GATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTGGGCCCACAAGGC 8460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8741 GATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTGGGCCCACAAGGC 8800

Qy       8461 CGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGGGCTGCCTGGGAA 8520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8801 CGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGGGCTGCCTGGGAA 8860

Qy       8521 ACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAGTACGCCCCGACC 8580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8861 ACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAGTACGCCCCGACC 8920

Qy       8581 ATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATGGCTCAAGACACG 8640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8921 ATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATGGCTCAAGACACG 8980

Qy       8641 CTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTGAGTCCCTTGGAC 8700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8981 CTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTGAGTCCCTTGGAC 9040

Qy       8701 CTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTGCACACATACACT 8760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9041 CTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTGCACACATACACT 9100

Qy       8761 CCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGA 8820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9101 CCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGA 9160

Qy       8821 GCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGTGGGGGGAGAGCG 8880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9161 GCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGTGGGGGGAGAGCG 9220

Qy       8881 GCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 8940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9221 GCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 9280

Qy       8941 TTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGCGCCGGCGGGGGC 9000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9281 TTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGCGCCGGCGGGGGC 9340

Qy       9001 GACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTTGGCCTACTCCTA 9060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9341 GACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTTGGCCTACTCCTA 9400

Qy       9061 CTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCGG 9099
             |||||||||||||||||||||||||||||||||||||||
Db       9401 CTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCGG 9439
```

```
RESULT 3
AAF23486
ID   AAF23486 standard; DNA; 9711 BP.
XX
AC   AAF23486;
XX
DT   21-MAR-2001  (first entry)
XX
DE   Infectious Hepatitis C virus 2a genotype.
XX
KW   GBV-B; hepatitis C virus; HCV; vaccine; ds.
XX
OS   Hepatitis C virus.
XX
PN   WO200075337-A1.
XX
PD   14-DEC-2000.
```

XX
PF 02-JUN-2000; 2000WO-US015293.
XX
PR 04-JUN-1999; 99US-0137694P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Bukh J, Yanagi M, Emerson SU, Purcell RH;
XX
DR WPI; 2001-091214/10.
XX
PT New infectious nucleic acids of the GB virus-B clone, useful for
PT indirectly studying the molecular properties of hepatitis C virus (HCV)
PT and in developing vaccines and therapeutics for HCV.
XX
PS Disclosure; Page 78-82; 96pp; English.
XX
CC The present invention relates to GB virus-B. The nucleic acid molecules
CC of the invention are useful for indirectly studying the molecular
CC properties of hepatitis C virus (HCV). The infectious nucleic acid
CC sequence of the GB virus-B clone and the HCV/GBV-B chimeras may be used
CC in the development of vaccines and therapeutics for HCV
XX
SQ Sequence 9711 BP; 1989 A; 2851 C; 2668 G; 2203 T; 0 U; 0 Other;

Query Match 100.0%; Score 9095.8; DB 4; Length 9711;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 9097; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 400

Qy         61 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 460

Qy        121 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 520

Qy        181 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 580

Qy        241 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 640

Qy        301 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        641 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 700

Qy        361 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        701 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 760

Qy        421 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        761 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 820

Qy        481 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        821 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 880

Qy        541 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        881 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 940

Qy        601 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        941 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 1000

Qy        661 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1001 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 1060

Qy        721 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1061 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 1120

Qy        781 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1121 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 1180

Qy        841 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1181 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 1240

Qy        901 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1241 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 1300

Qy        961 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1301 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1360

Qy       1021 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1361 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1420

Qy       1081 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1421 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1480

Qy       1141 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1481 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1540

Qy       1201 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1541 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1600

Qy       1261 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1601 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1660

Qy       1321 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1661 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1720

Qy       1381 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1721 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1780

Qy       1441 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1781 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1840

Qy       1501 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1841 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1900

Qy       1561 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1901 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1960

Qy       1621 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1961 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 2020

Qy       1681 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2021 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 2080

Qy       1741 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 1800
```

```
Db      2081 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 2140
Qy      1801 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 1860
Db      2141 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 2200
Qy      1861 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 1920
Db      2201 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 2260
Qy      1921 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 1980
Db      2261 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAAATTGGAGGAC 2320
Qy      1981 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2040
Db      2321 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2380
Qy      2041 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2100
Db      2381 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2440
Qy      2101 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2160
Db      2441 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2500
Qy      2161 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2220
Db      2501 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2560
Qy      2221 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2280
Db      2561 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2620
Qy      2281 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATATTTTGTCATCTTTTTCGTGGCTGCTTGG 2340
Db      2621 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATATTTTGTCATCTTTTTCGTGGCTGCTTGG 2680
Qy      2341 TACATCAAGGGTCGGGTAGTCCCCTTAGCTACCTATTCCCTCACTGGCCTGTGGTCCTTT 2400
Db      2681 TACATCAAGGGTCGGGTAGTCCCCTTAGCTACCTATTCCCTCACTGGCCTGTGGTCCTTT 2740
Qy      2401 AGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTGCATGGC 2460
Db      2741 AGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTGCATGGC 2800
Qy      2461 CAGATAGGAGCGGCTCTGCTGGTAATGATCACTCTCTTTACTCTCACCCCCGGGTATAAG 2520
Db      2801 CAGATAGGAGCGGCTCTGCTGGTAATGATCACTCTCTTTACTCTCACCCCCGGGTATAAG 2860
Qy      2521 ACCCTTCTCAGCCGGTTTTTGTGGTGGTTGTGCTATCTTCTGACCCTGGGGGAAGCTATG 2580
Db      2861 ACCCTTCTCAGCCGGTTTTTGTGGTGGTTGTGCTATCTTCTGACCCTGGGGGAAGCTATG 2920
Qy      2581 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGCATCATATGGGCC 2640
Db      2921 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGCATCATATGGGCC 2980
Qy      2641 GTCGCCATATTCTACCCAGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 2700
Db      2981 GTCGCCATATTCTACCCAGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 3040
Qy      2701 GGGCCTGCTTACCTCCTAAAAGGTGCTTTGACGCGCGTGCCGTACTTCGTCAGGGCTCAC 2760
Db      3041 GGGCCTGCTTACCTCCTAAAAGGTGCTTTGACGCGCGTGCAGTACTTCGTCAGGGCTCAC 3100
Qy      2761 GCTCTACTGAGGATGTGCACCATGGCAAGGCATCTCGCGGGGGGCAGGTACGTCCAGATG 2820
Db      3101 GCTCTACTGAGGATGTGCACCATGGCAAGGCATCTCGCGGGGGGCAGGTACGTCCAGATG 3160
Qy      2821 GCGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 2880
Db      3161 GCGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 3220
Qy      2881 TCGGATTGGGCTGCTAGTGGCCTGCGGGACCTGGCGGTCGCCGTTGAGCCTATCATCTTC 2940
Db      3221 TCGGATTGGGCTGCTAGTGGCCTGCGGGACCTGGCGGTCGCCGTTGAGCCTATCATCTTC 3280
Qy      2941 AGTCCGATGGAGAAGAAAGTCATTGTCTGGGGAGCGGAGACAGCTGCTTGTGGGGACATT 3000
Db      3281 AGTCCGATGGAGAAGAAAGTCATTGTCTGGGGAGCGGAGACAGCTGCTTGTGGGGACATT 3340
Qy      3001 TTACACGGACTTCCCGTGTCCGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3060
Db      3341 TTACACGGACTTCCCGTGTCCGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3400
Qy      3061 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3120
Db      3401 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3460
Qy      3121 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGGCGCGACAAGACAGAACAGGCT 3180
Db      3461 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGGCGCGACAAGACAGAACAGGCT 3520
Qy      3181 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3240
Db      3521 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3580
Qy      3241 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3300
Db      3581 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3640
Qy      3301 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGG 3360
Db      3641 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGG 3700
Qy      3361 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3420
Db      3701 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3760
Qy      3421 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3480
Db      3761 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3820
Qy      3481 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3540
Db      3821 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3880
Qy      3541 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3600
Db      3881 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3940
Qy      3601 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 3660
Db      3941 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 4000
Qy      3661 CCTGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 3720
Db      4001 CCTGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 4060
Qy      3721 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 3780
Db      4061 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 4120
Qy      3781 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 3840
Db      4121 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 4180
Qy      3841 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 3900
Db      4181 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 4240
Qy      3901 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 3960
Db      4241 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 4300
```

```
Qy      3961 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4301 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4360

Qy      4021 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4361 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4420

Qy      4081 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4421 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4480

Qy      4141 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4481 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4540

Qy      4201 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4541 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4600

Qy      4261 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4601 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4660

Qy      4321 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4661 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4720

Qy      4381 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4721 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4780

Qy      4441 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4781 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4840

Qy      4501 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4841 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4900

Qy      4561 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4901 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4960

Qy      4621 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 4680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4961 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 5020

Qy      4681 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 4740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5021 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 5080

Qy      4741 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 4800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5081 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 5140

Qy      4801 GCCAAAGCCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 4860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5141 GCCAAAGCCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 5200

Qy      4861 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 4920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5201 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 5260

Qy      4921 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 4980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5261 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 5320

Qy      4981 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5321 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5380

Qy      5041 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5381 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5440

Qy      5101 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5441 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5500

Qy      5161 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5501 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5560

Qy      5221 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5561 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5620

Qy      5281 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5621 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5680

Qy      5341 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5681 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5740

Qy      5401 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5741 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5800

Qy      5461 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5801 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5860

Qy      5521 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5861 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5920

Qy      5581 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5921 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5980

Qy      5641 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 5700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5981 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 6040

Qy      5701 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 5760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6041 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 6100

Qy      5761 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 5820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6101 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 6160

Qy      5821 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 5880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6161 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 6220

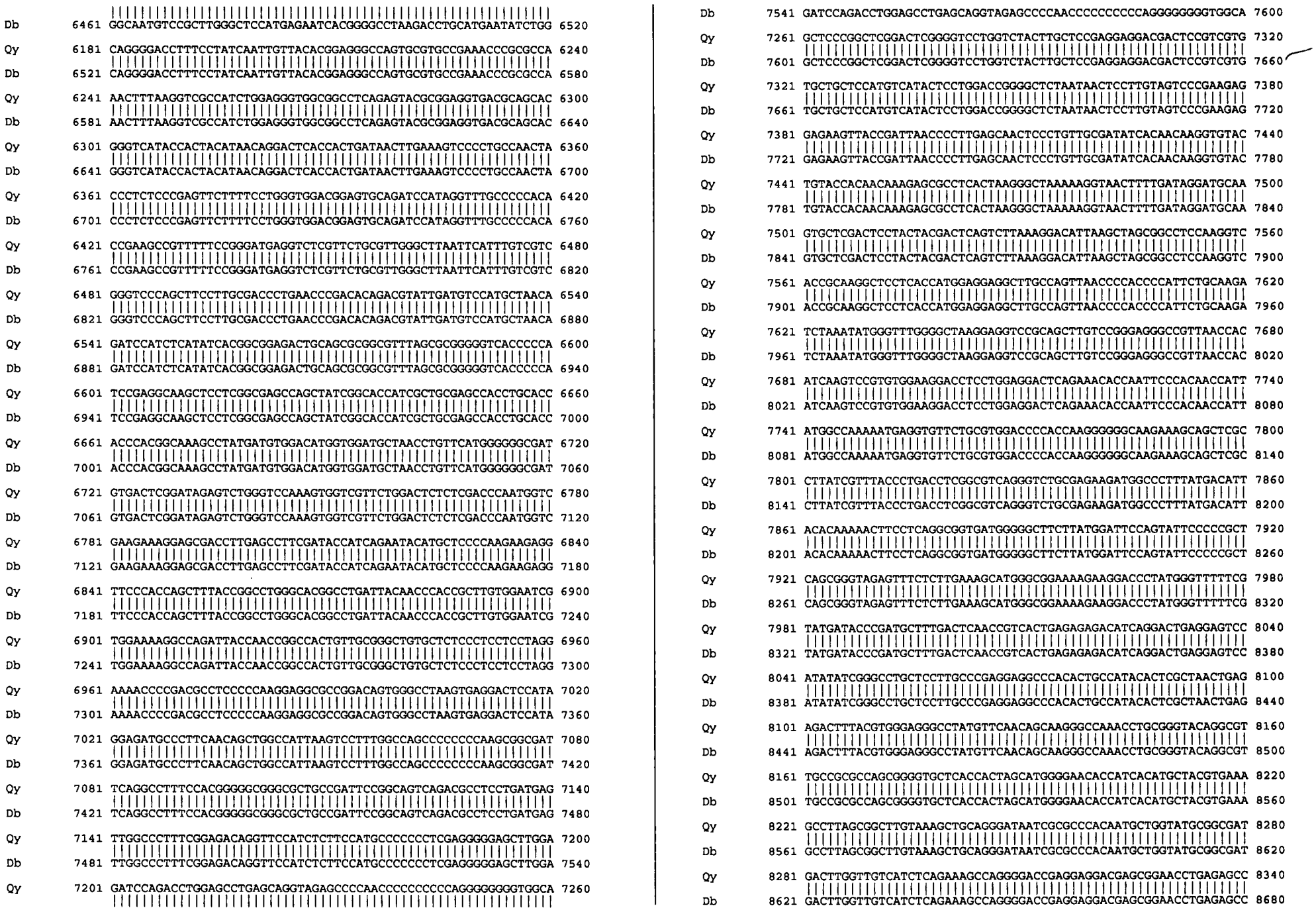
Qy      5881 CTCAGAAGACTCCACAACTGGATTACTGAGGACTGCCCCATCCCATGCGGCGGCTCGTGG 5940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6221 CTCAGAAGACTCCACAACTGGATTACTGAGGACTGCCCCATCCCATGCGGCGGCTCGTGG 6280

Qy      5941 CTCCGCGATGTGTGGGACTGGGTTTGCACCATCCTAACAGACTTTAAAAATTGGCTGACC 6000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6281 CTCCGCGATGTGTGGGACTGGGTTTGCACCATCCTAACAGACTTTAAAAATTGGCTGACC 6340

Qy      6001 TCCAAATTATTCCCAAAGATGCCCGGCCTCCCCTTTGTCTCCTGTCAAAAGGGGTACAAG 6060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6341 TCCAAATTATTCCCAAAGATGCCCGGCCTCCCCTTTGTCTCCTGTCAAAAGGGGTACAAG 6400

Qy      6061 GGCGTGTGGGCCGGCACTGGCATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6401 GGCGTGTGGGCCGGCACTGGCATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6460

Qy      6121 GGCAATGTCCGCTTGGGCTCCATGAGAATCACGGGGCCTAAGACCTGCATGAATATCTGG 6180
```



```
Qy       8341 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCCAGACCGGAGTAT 8400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8681 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCCAGACCGGAGTAT 8740

Qy       8401 GATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTGGGCCCACAAGGC 8460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8741 GATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTGGGCCCACAAGGC 8800

Qy       8461 CGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGGGCTGCCTGGGAA 8520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8801 CGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGGGCTGCCTGGGAA 8860

Qy       8521 ACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAGTACGCCCCGACC 8580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8861 ACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAGTACGCCCCGACC 8920

Qy       8581 ATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATGGCTCAAGACACG 8640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8921 ATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATGGCTCAAGACACG 8980

Qy       8641 CTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTGAGTCCCTTGGAC 8700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8981 CTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTGAGTCCCTTGGAC 9040

Qy       8701 CTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTGCACACATACACT 8760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9041 CTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTGCACACATACACT 9100

Qy       8761 CCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGA 8820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9101 CCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGA 9160

Qy       8821 GCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGTGGGGGGAGAGCG 8880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9161 GCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGTGGGGGGAGAGCG 9220

Qy       8881 GCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 8940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9221 GCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 9280

Qy       8941 TTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGCGCCGGCGGGGGC 9000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9281 TTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGCGCCGGCGGGGGC 9340

Qy       9001 GACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTTGGCCTACTCCTA 9060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9341 GACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTTGGCCTACTCCTA 9400

Qy       9061 CTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCGG 9099
             |||||||||||||||||||||||||||||||||||||||
Db       9401 CTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCGG 9439
```

```
RESULT 4
AAQ38218
ID   AAQ38218 standard; cDNA to mRNA; 9589 BP.
XX
AC   AAQ38218;
XX
DT   27-AUG-2003  (revised)
DT   25-MAR-2003  (revised)
DT   01-JUL-1993  (first entry)
XX
DE   NANBH virus strain HC-J6 genome cDNA sequence.
XX
KW   Non A non B hepatitis virus; amplification; HC-J1; HC-J8; plasma; ss.
XX
OS   Non-A.
OS   non-B hepatitis virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             341. .9442
FT                   /*tag= a
XX
PN   EP532167-A2.
XX
PD   17-MAR-1993.
XX
PF   30-JUL-1992;   92EP-00306952.
XX
PR   09-AUG-1991;   91JP-00287402.
PR   05-DEC-1991;   91JP-00360441.
XX
PA   (IMMO ) IMMUNO JAPAN INC.
XX
PI   Okamoto H,  Nakamura T;
XX
DR   WPI; 1993-087166/11.
DR   N-PSDB; AAR33538.
XX
PT   Polynucleotide(s), polypeptide(s) and antibodies of NANBH virus - useful
PT   for detecting NANBH, as a vaccine and for screening blood samples.
XX
PS   Claim 2;. Page 27-32; 93pp; English.
XX
CC   RNA was isolated from the plasma of human patients positive for NANBH
CC   virus (strain HC-J6) and was subjected to reverse transcription to
CC   produce cDNA. The resulting cDNA was amplified by PCR, and nucleic acid
CC   sequences determined by analysis of both clones from the cDNA library and
CC   clones obtd. by PCR amplification (36 clones in total). The NANBH HC-J6
CC   genome was found to contain an open reading frame encoding a polypeptide
CC   precursor of 3033 amino acid residues. See also AAQ38172-221. (Updated on
CC   25-MAR-2003 to correct PN field.) (Updated on 27-AUG-2003 to correct OS
CC   field.)
XX
SQ   Sequence 9589 BP; 1968 A; 2820 C; 2635 G; 2166 T; 0 U; 0 Other;

  Query Match             93.4%;  Score 8494.2;  DB 2;  Length 9589;
  Best Local Similarity   95.8%;  Pred. No. 0;
  Matches 8721;  Conservative    0;  Mismatches  378;  Indels    0;  Gaps    0;

Qy          1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 400

Qy         61 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 460

Qy        121 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 520

Qy        181 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 240
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        521 AGGCGCCAGCCCATCCCTAAGGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 580

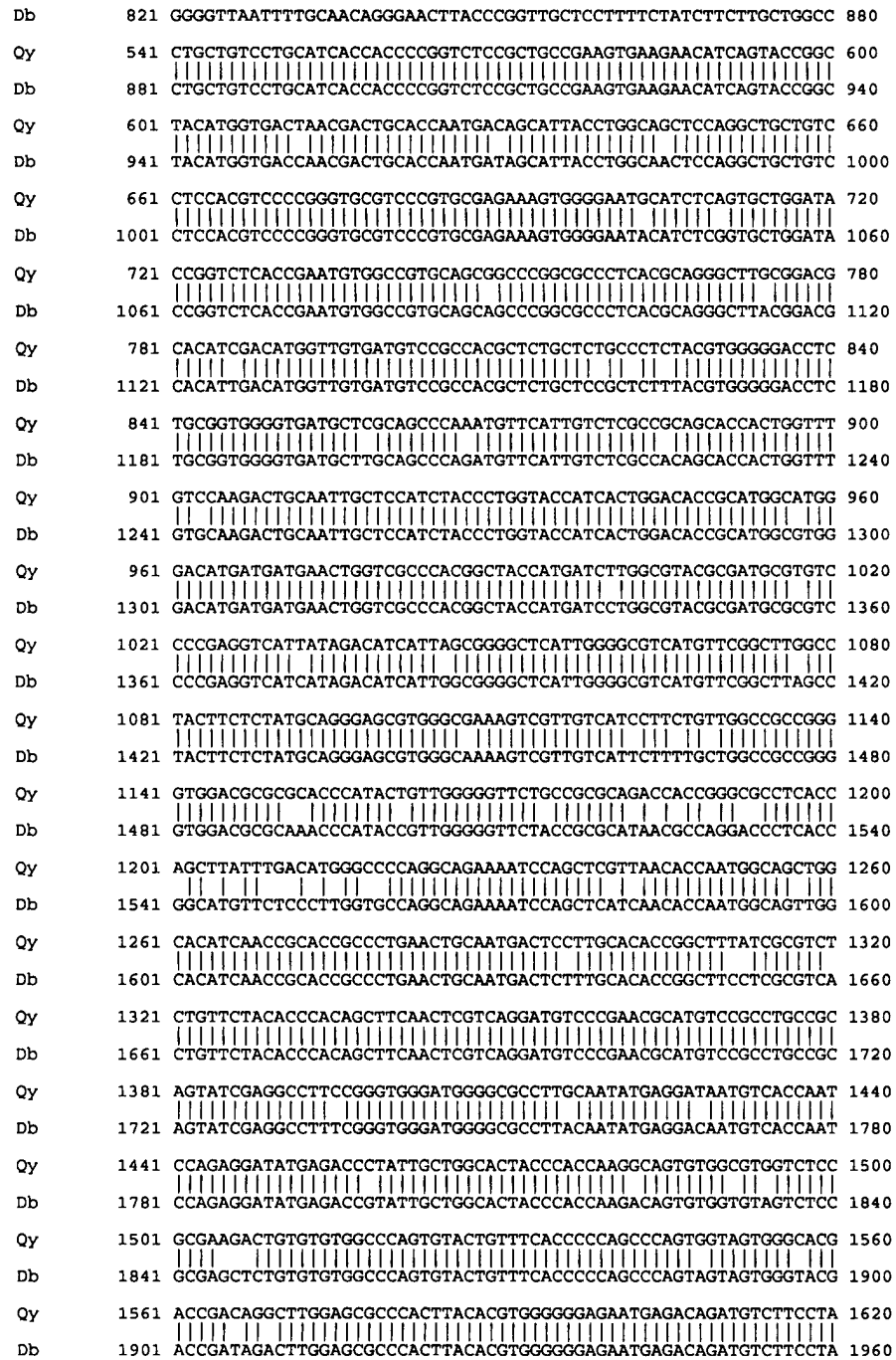
Qy        241 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 640

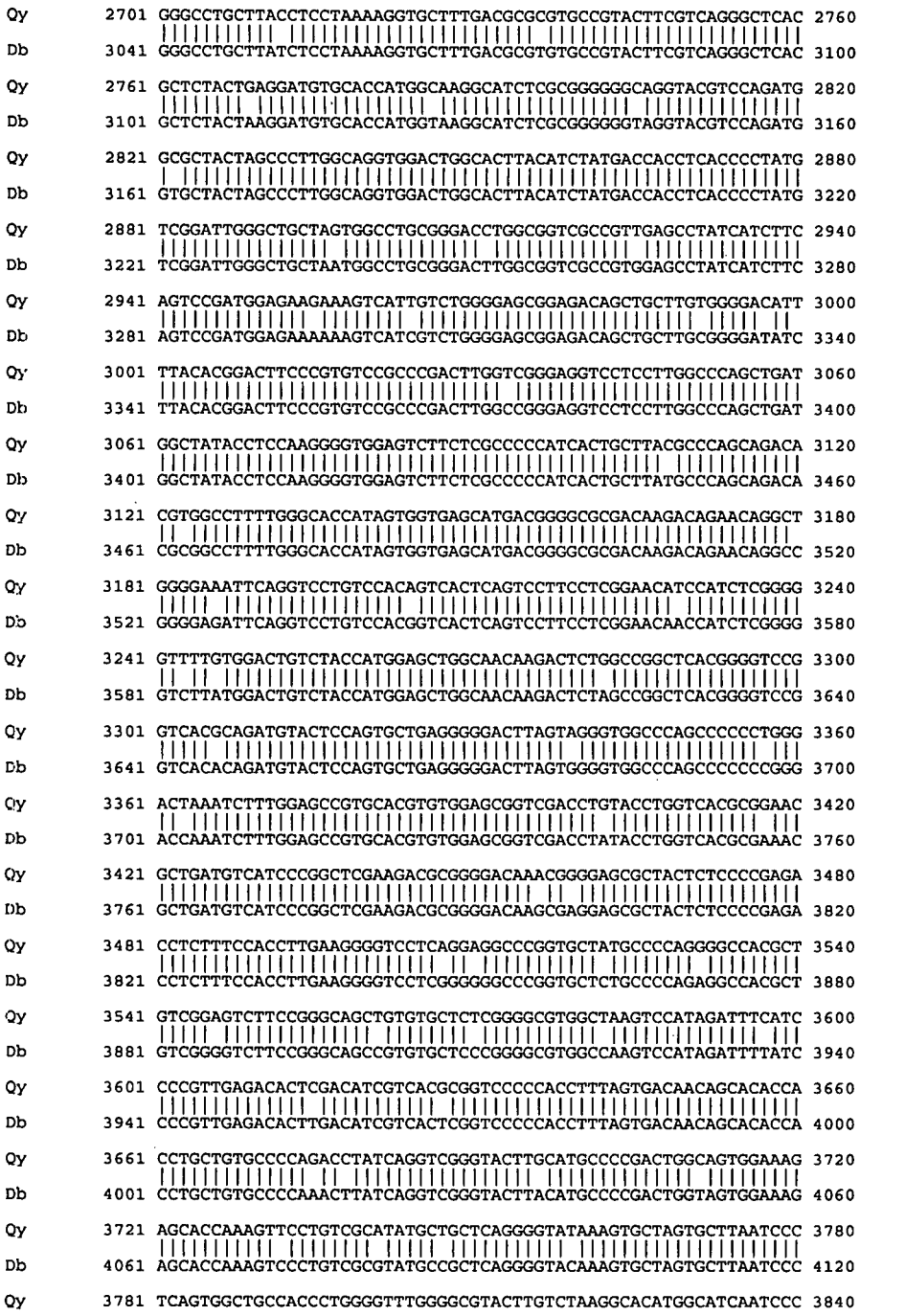
Qy        301 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 360
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        641 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCCCGCAACGTGGGT 700

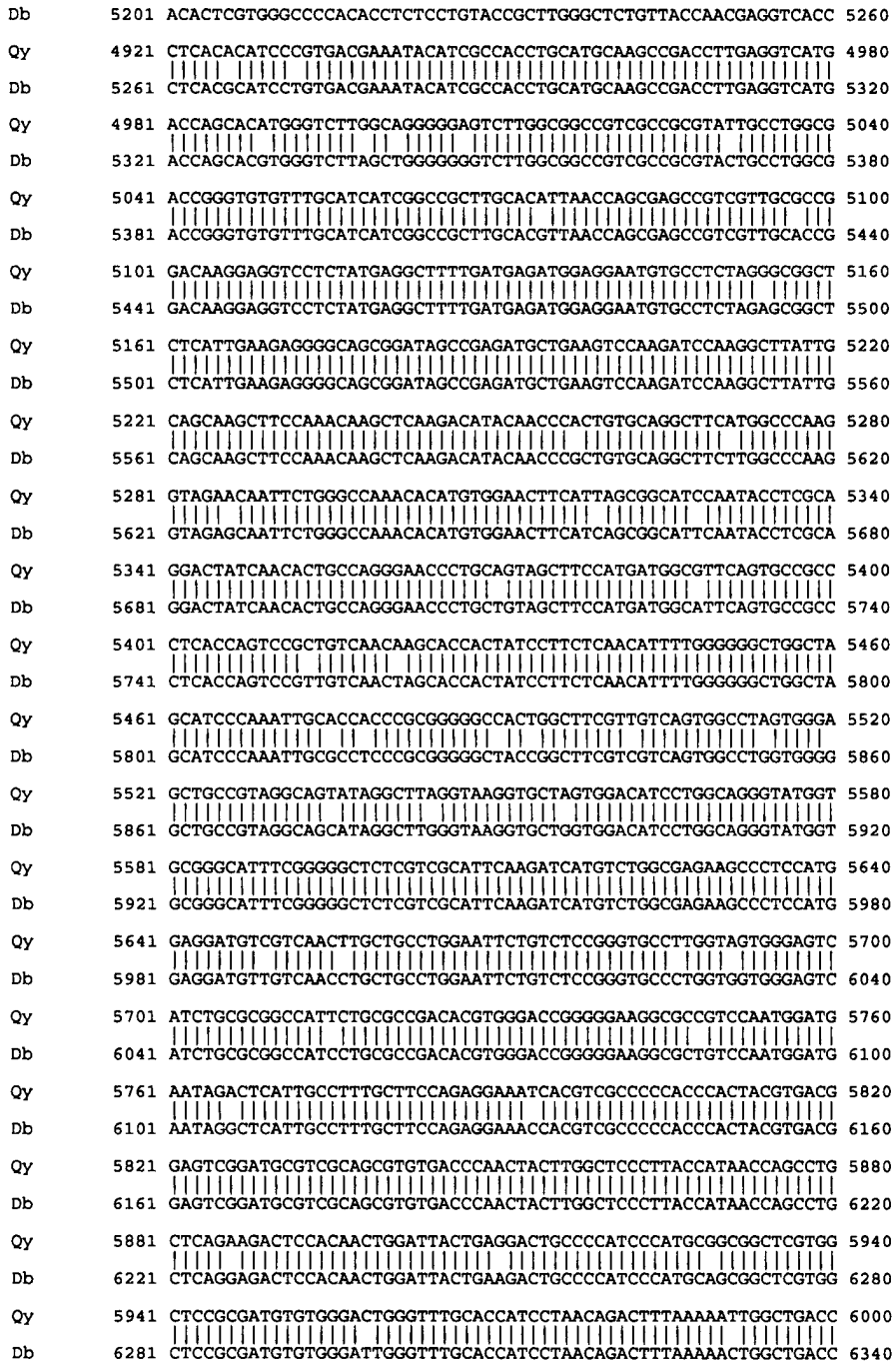
Qy        361 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        701 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTA 760

Qy        421 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        761 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 820

Qy        481 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```


Db 821 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 880
Qy 541 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 600
Db 881 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 940
Qy 601 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 660
Db 941 TACATGGTGACCAACGACTGCACCAATGATAGCATTACCTGGCAACTCCAGGCTGCTGTC 1000
Qy 661 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 720
Db 1001 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATACATCTCGGTGCTGGATA 1060
Qy 721 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 780
Db 1061 CCGGTCTCACCGAATGTGGCCGTGCAGCAGCCCGGCGCCCTCACGCAGGGCTTACGGACG 1120
Qy 781 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 840
Db 1121 CACATTGACATGGTTGTGATGTCCGCCACGCTCTGCTCCGCTCTTTACGTGGGGGACCTC 1180
Qy 841 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 900
Db 1181 TGCGGTGGGGTGATGCTTGCAGCCCAGATGTTCATTGTCTCGCCACAGCACCACTGGTTT 1240
Qy 901 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 960
Db 1241 GTGCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCGTGG 1300
Qy 961 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1020
Db 1301 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCCTGGCGTACGCGATGCGCGTC 1360
Qy 1021 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1080
Db 1361 CCCGAGGTCATCATAGACATCATTGGCGGGGCTCATTGGGGCGTCATGTTCGGCTTAGCC 1420
Qy 1081 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1140
Db 1421 TACTTCTCTATGCAGGGAGCGTGGGCAAAAGTCGTTGTCATTCTTTTGCTGGCCGCCGGG 1480
Qy 1141 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1200
Db 1481 GTGGACGCGCAAACCCATACCGTTGGGGGTTCTACCGCGCATAACGCCAGGACCCTCACC 1540
Qy 1201 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1260
Db 1541 GGCATGTTCTCCCTTGGTGCCAGGCAGAAAATCCAGCTCATCAACACCAATGGCAGTTGG 1600
Qy 1261 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1320
Db 1601 CACATCAACCGCACCGCCCTGAACTGCAATGACTCTTTGCACACCGGCTTCCTCGCGTCA 1660
Qy 1321 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1380
Db 1661 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1720
Qy 1381 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1440
Db 1721 AGTATCGAGGCCTTTCGGGTGGGATGGGGCGCCTTACAATATGAGGACAATGTCACCAAT 1780
Qy 1441 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1500
Db 1781 CCAGAGGATATGAGACCGTATTGCTGGCACTACCCACCAAGACAGTGTGGTGTAGTCTCC 1840
Qy 1501 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1560
Db 1841 GCGAGCTCTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTAGTAGTGGGTACG 1900
Qy 1561 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1620
Db 1901 ACCGATAGACTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1960
Qy 1621 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 1680
Db 1961 TTGAACAGCACTCGACCACCGCAGGGGTCATGGTTCGGCTGCACGTGGATGAACTCCACT 2020
Qy 1681 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 1740
Db 2021 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGCATTAGAGCTGACTTCAATGCCAGC 2080
Qy 1741 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 1800
Db 2081 ATGGACTTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACCTACATCAAA 2140
Qy 1801 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 1860
Db 2141 TGTGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 2200
Qy 1861 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 1920
Db 2201 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGGGGGGTC 2260
Qy 1921 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 1980
Db 2261 GAGCACAGGCTCACGGCTGCGTGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 2320
Qy 1981 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2040
Db 2321 AGAGACAGAAGTCAACTGTCTCCTTTGCTGCACTCCACCACGGAGTGGGCCATTTTACCT 2380
Qy 2041 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2100
Db 2381 TGCACTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2440
Qy 2101 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2160
Db 2441 GTGGACGTGCAATTCATGTATGGCCTATCACCTGCTCTCACAAAATACATCGTCCGATGG 2500
Qy 2161 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2220
Db 2501 GAGTGGGTAGTACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2560
Qy 2221 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2280
Db 2561 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGTTGGTCGTCTTGCACGCT 2620
Qy 2281 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATATTTTGTCATCTTTTTCGTGGCTGCTTGG 2340
Db 2621 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATACTTTGTCATCTTTTTCGTGGCTGCTTGG 2680
Qy 2341 TACATCAAGGGTCGGGTAGTCCCCTTAGCTACCTATTCCCTCACTGGCCTGTGGTCCTTT 2400
Db 2681 TACATCAAGGGTCGGGTAGTCCCCTTGGCTACTTATTCCCTCACTGGCCTATGGTCCTTT 2740
Qy 2401 AGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTGCATGGC 2460
Db 2741 GGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTACATGGT 2800
Qy 2461 CAGATAGGAGCGGCTCTGCTGGTAATGATCACTCTCTTTACTCTCACCCCCGGGTATAAG 2520
Db 2801 CAGATAGGAGCAGCTCTGTTGGTACTGATCACTCTCTTTACACTCACCCCCGGGTATAAG 2860
Qy 2521 ACCCTTCTCAGCCGGTTTTTGTGGTGGTTGTGCTATCTTCTGACCCTGGGGGAAGCTATG 2580
Db 2861 ACCCTTCTCAGCCGGTTTCTGTGGTGGTTGTGCTATCTTCTGACCCTGGCGGAAGCTATG 2920
Qy 2581 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGCATCATATGGGCC 2640
Db 2921 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGGATCATATGGGCC 2980
Qy 2641 GTCGCCATATTCTACCCAGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 2700
Db 2981 GTCGCCATATTCTGCCCGGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 3040


Qy 2701 GGGCCTGCTTACCTCCTAAAAGGTGCTTTGACGCGCGTGCCGTACTTCGTCAGGGCTCAC 2760
Db 3041 GGGCCTGCTTATCTCCTAAAAGGTGCTTTGACGCGTGTGCCGTACTTCGTCAGGGCTCAC 3100
Qy 2761 GCTCTACTGAGGATGTGCACCATGGCAAGGCATCTCGCGGGGGGCAGGTACGTCCAGATG 2820
Db 3101 GCTCTACTAAGGATGTGCACCATGGTAAGGCATCTCGCGGGGGGTAGGTACGTCCAGATG 3160
Qy 2821 GCGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 2880
Db 3161 GTGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 3220
Qy 2881 TCGGATTGGGCTGCTAGTGGCCTGCGGGACCTGGCGGTCGCCGTTGAGCCTATCATCTTC 2940
Db 3221 TCGGATTGGGCTGCTAATGGCCTGCGGGACTTGGCGGTCGCCGTGGAGCCTATCATCTTC 3280
Qy 2941 AGTCCGATGGAGAAGAAAGTCATTGTCTGGGGAGCGGAGACAGCTGCTTGTGGGGACATT 3000
Db 3281 AGTCCGATGGAGAAAAAAGTCATCGTCTGGGGAGCGGAGACAGCTGCTTGCGGGGATATC 3340
Qy 3001 TTACACGGACTTCCCGTGTCCGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3060
Db 3341 TTACACGGACTTCCCGTGTCCGCCCGACTTGGCCGGGAGGTCCTCCTTGGCCCAGCTGAT 3400
Qy 3061 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3120
Db 3401 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTATGCCCAGCAGACA 3460
Qy 3121 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCT 3180
Db 3461 CGCGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCC 3520
Qy 3181 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3240
Db 3521 GGGGAGATTCAGGTCCTGTCCACGGTCACTCAGTCCTTCCTCGGAACAACCATCTCGGGG 3580
Qy 3241 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3300
Db 3581 GTCTTATGGACTGTCTACCATGGAGCTGGCAACAAGACTCTAGCCGGCTCACGGGGTCCG 3640
Qy 3301 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGG 3360
Db 3641 GTCACACAGATGTACTCCAGTGCTGAGGGGGACTTAGTGGGGTGGCCCAGCCCCCCCGGG 3700
Qy 3361 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3420
Db 3701 ACCAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTATACCTGGTCACGCGAAAC 3760
Qy 3421 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3480
Db 3761 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAGCGAGGAGCGCTACTCTCCCCGAGA 3820
Qy 3481 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3540
Db 3821 CCTCTTTCCACCTTGAAGGGGTCCTCGGGGGGCCCGGTGCTCTGCCCCAGAGGCCACGCT 3880
Qy 3541 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3600
Db 3881 GTCGGGGTCTTCCGGGCAGCCGTGTGCTCCCGGGGCGTGGCCAAGTCCATAGATTTTATC 3940
Qy 3601 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 3660
Db 3941 CCCGTTGAGACACTTGACATCGTCACTCGGTCCCCCACCTTTAGTGACAACAGCACACCA 4000
Qy 3661 CCTGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 3720
Db 4001 CCTGCTGTGCCCCAAACTTATCAGGTCGGGTACTTACATGCCCCGACTGGTAGTGGAAAG 4060
Qy 3721 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 3780
Db 4061 AGCACCAAAGTCCCTGTCGCGTATGCCGCTCAGGGGTACAAAGTGCTAGTGCTTAATCCC 4120
Qy 3781 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 3840
Db 4121 TCGGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCCAAGGCACATGGCATCAATCCC 4180
Qy 3841 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 3900
Db 4181 AACATTAGGACTGGGGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 4240
Qy 3901 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 3960
Db 4241 GGCAAATTCCTCGCCGATGGGGGCTGCGCAGGCGGCGCCTATGACATCATCATATGCGAT 4300
Qy 3961 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4020
Db 4301 GAATGCCATGCCGTGGACTCTACCACCATTCTCGGCATCGGAACAGTCCTCGATCAAGCA 4360
Qy 4021 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4080
Db 4361 GAGACAGCCGGGGTCAGGCTAACTGTACTGGCTACGGCTACGCCCCCCGGGTCAGTGACA 4420
Qy 4081 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4140
Db 4421 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTCGGGCAGGAGGGTGAGATCCCCTTCTAT 4480
Qy 4141 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4200
Db 4481 GGGAGGGCGATTCCCCTGTCATACATCAAGGGAGGAAGACACTTGATCTTCTGCCACTCA 4540
Qy 4201 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4260
Db 4541 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACGCAGTGGCA 4600
Qy 4261 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4320
Db 4601 TACTACAGAGGGCTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4660
Qy 4321 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4380
Db 4661 ACCGACGCCCTCATGACGGGGTTTACTGGAGACTTTGACTCCGTGATCGACTGCAACGTA 4720
Qy 4381 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4440
Db 4721 GCGGTCACTCAAGTTGTAGACTTCAGCTTGGACCCCACATTCACCATAACCACACAGACT 4780
Qy 4441 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4500
Db 4781 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGCCGCACGGGCAGGGGAAGACTG 4840
Qy 4501 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4560
Db 4841 GGTATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4900
Qy 4561 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4620
Db 4901 CTCTGCGAGTGCTACGATGCAGGGGCCGCATGGTATGAGCTCACACCAGCGGAGACCACC 4960
Qy 4621 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 4680
Db 4961 GTCAGGCTCAGAGCATATTTCAACACACCTGGTTTGCCTGTGTGCCAAGACCATCTTGAG 5020
Qy 4681 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 4740
Db 5021 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 5080
Qy 4741 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 4800
Db 5081 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 5140
Qy 4801 GCCAAAGCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 4860
Db 5141 GCCAAAGCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 5200
Qy 4861 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 4920


Db 5201 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 5260
Qy 4921 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 4980
Db 5261 CTCACGCATCCTGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 5320
Qy 4981 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5040
Db 5321 ACCAGCACGTGGGTCTTAGCTGGGGGGGTCTTGGCGGCCGTCGCCGCGTACTGCCTGGCG 5380
Qy 5041 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5100
Db 5381 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACGTTAACCAGCGAGCCGTCGTTGCACCG 5440
Qy 5101 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5160
Db 5441 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGAGCGGCT 5500
Qy 5161 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5220
Db 5501 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5560
Qy 5221 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5280
Db 5561 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCGCTGTGCAGGCTTCTTGGCCCAAG 5620
Qy 5281 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5340
Db 5621 GTAGAGCAATTCTGGGCCAAACACATGTGGAACTTCATCAGCGGCATTCAATACCTCGCA 5680
Qy 5341 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5400
Db 5681 GGACTATCAACACTGCCAGGGAACCCTGCTGTAGCTTCCATGATGGCATTCAGTGCCGCC 5740
Qy 5401 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5460
Db 5741 CTCACCAGTCCGTTGTCAACTAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5800
Qy 5461 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5520
Db 5801 GCATCCCAAATTGCGCCTCCCGCGGGGGCTACCGGCTTCGTCGTCAGTGGCCTGGTGGGG 5860
Qy 5521 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5580
Db 5861 GCTGCCGTAGGCAGCATAGGCTTGGGTAAGGTGCTGGTGGACATCCTGGCAGGGTATGGT 5920
Qy 5581 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5640
Db 5921 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5980
Qy 5641 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 5700
Db 5981 GAGGATGTTGTCAACCTGCTGCCTGGAATTCTGTCTCCGGGTGCCCTGGTGGTGGGAGTC 6040
Qy 5701 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 5760
Db 6041 ATCTGCGCGGCCATCCTGCGCCGACACGTGGGACCGGGGGAAGGCGCTGTCCAATGGATG 6100
Qy 5761 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 5820
Db 6101 AATAGGCTCATTGCCTTTGCTTCCAGAGGAAACCACGTCGCCCCCACCCACTACGTGACG 6160
Qy 5821 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 5880
Db 6161 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 6220
Qy 5881 CTCAGAAGACTCCACAACTGGATTACTGAGGACTGCCCCATCCCATGCGGCGGCTCGTGG 5940
Db 6221 CTCAGGAGACTCCACAACTGGATTACTGAAGACTGCCCCATCCCATGCAGCGGCTCGTGG 6280
Qy 5941 CTCCGCGATGTGTGGGACTGGGTTTGCACCATCCTAACAGACTTTAAAAATTGGCTGACC 6000
Db 6281 CTCCGCGATGTGTGGGATTGGGTTTGCACCATCCTAACAGACTTTAAAAACTGGCTGACC 6340
Qy 6001 TCCAAATTATTCCCAAAGATGCCCGGCCTCCCCTTTGTCTCCTGTCAAAAGGGGTACAAG 6060
Db 6341 TCCAAATTGTTCCCAAAGATGCCTGGTCTCCCCTTTATCTCTTGTCAAAAGGGGTACAAG 6400
Qy 6061 GGCGTGTGGGCCGGCACTGGCATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6120
Db 6401 GGCGTGTGGGCTGGCACTGGTATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6460
Qy 6121 GGCAATGTCCGCTTGGGCTCCATGAGAATCACGGGGGCCTAAGACCTGCATGAATATCTGG 6180
Db 6461 GGCAATGTCCGCCTGGGCTCCATGAGAATTACGGGGCCCAAAACCTGCATGAATATCTGG 6520
Qy 6181 CAGGGGACCTTTCCTATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCGCCA 6240
Db 6521 CAGGGGACCTTTCCCATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCACCA 6580
Qy 6241 AACTTTAAGGTCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6300
Db 6581 AACTTTAAGATCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6640
Qy 6301 GGGTCATACCACTACATAACAGGACTCACCACTGATAACTTGAAAGTCCCCTGCCAACTA 6360
Db 6641 GGGTCATACCACTACATAACAGGACTTACCACTGATAACTTGAAAGTTCCTTGCCAACTA 6700
Qy 6361 CCCTCTCCCGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCACA 6420
Db 6701 CCTTCTCCAGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCATA 6760
Qy 6421 CCGAAGCCGTTTTTCCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6480
Db 6761 CCGAAGCCGTTTTTTCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6820
Qy 6481 GGGTCCCAGCTTCCTTGCGACCCTGAACCCGACACAGACGTATTGATGTCCATGCTAACA 6540
Db 6821 GGGTCTCAGCTCCCTTGCGATCCTGAACCTGACACAGACGTATTGACGTCCATGCTAACA 6880
Qy 6541 GATCCATCTCATATCACGGCGGAGACTGCAGCGCGGCGTTTAGCGCGGGGGTCACCCCCA 6600
Db 6881 GACCCATCCCATATCACGGCGGAGACTGCAGCGCGGCGTTTGGCACGGGGGTCACCCCCG 6940
Qy 6601 TCCGAGGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 6660
Db 6941 TCCGAGGCAAGCTCCTCAGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 7000
Qy 6661 ACCCACGGCAAAGCCTATGATGTGGACATGGTGGATGCTAACCTGTTCATGGGGGGCGAT 6720
Db 7001 ACCCACGGCAAGGCCTATGATGTGGACATGGTGGATGCCAACCTGTTCATGGGGGGCGAT 7060
Qy 6721 GTGACTCGGATAGAGTCTGGGTCCAAAGTGGTCGTTCTGGACTCTCTCGACCCAATGGTC 6780
Db 7061 GTGACCCGGATAGAGTCTGAGTCCAAAGTGGTCGTTCTGGACTCTCTCGACCCAATGGTC 7120
Qy 6781 GAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCAGAATACATGCTCCCCAAGAAGAGG 6840
Db 7121 GAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCGGAATATATGCTCCCCAAGAAGAGA 7180
Qy 6841 TTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCGCTTGTGGAATCG 6900
Db 7181 TTCCCACCAGCCTTACCGGCTTGGGCACGGCCTGATTACAACCCACCGCTTGTGGAATCG 7240
Qy 6901 TGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTCCCTCCTCCTAGG 6960
Db 7241 TGGAAGAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGCGCTCTCCCCCCCCCTAAG 7300
Qy 6961 AAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGTGAGGACTCCATA 7020
Db 7301 AAAACCCCGACGCCTCCCCCAAGGAGACGCCGGACAGTGGGTCTGAGTGAGAGCTCCATA 7360
Qy 7021 GGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCCCCCAAGCGGCGAT 7080
Db 7361 GCAGATGCCCTACAACAGCTGGCCATCAAGTCCTTTGGCCAGCCCCCCCCAAGCGGCGAT 7420



RESULT 5
ABK88904
ID ABK88904 standard; cDNA; 9678 BP.
XX
AC ABK88904;
XX
DT 07-OCT-2002 (first entry)
XX
DE Human HCV-related polypeptide encoding cDNA.

```
XX
KW   Hepatitis C virus; human; gene; ss; virucide; gene therapy; HCV;
KW   fulminant hepatitis C.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             341. .9442
FT                   /*tag= a
FT                   /product= "Human HCV-related protein"
XX
PN   JP2002171978-A.
XX
PD   18-JUN-2002.
XX
PF   01-DEC-2000; 2000JP-00367365.
XX
PR   01-DEC-2000; 2000JP-00367365.
XX
PA   (TOKR-) ZH TOKYOTO RINSHO IGAKU SOGO KENKYUSHO.
PA   (TORA ) TORAY IND INC.
XX
DR   WPI; 2002-569884/61.
DR   P-PSDB; ABG30688.
XX
PT   A gene of a fulminant hepatitis C virus strain and the encoded
PT   polypeptide useful in gene therapy to treat hepatitis C.
XX
PS   Claim 5; Page 13-25; 36pp; Japanese.
XX
CC   The invention relates to a human polypeptide related to hepatitis C virus
CC   (HCV), and the polynucleotide encoding it. The polypeptide can be used
CC   for the development of gene therapy on fulminant hepatitis C. This
CC   sequence represents cDNA encoding a human HCV-related polypeptide
XX
SQ   Sequence 9678 BP; 1948 A; 2915 C; 2723 G; 2092 T; 0 U; 0 Other;

  Query Match             82.6%;  Score 7513.4;  DB 6;  Length 9678;
  Best Local Similarity   89.1%;  Pred. No. 0;
  Matches 8108;  Conservative    0;  Mismatches  991;  Indels    0;  Gaps    0;

Qy          1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        341 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCAGAA 400

Qy         61 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 120
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 GACGTTAAGTTCCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 460

Qy        121 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 180
              |||||||||||||||||||| |||||||||| ||||||||||||||||||||||||||||
Db        461 GGCCCCAGGTTGGGTGTGCGCACGACAAGGAAAACTTCGGAGCGGTCCCAGCCACGTGGG 520

Qy        181 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 240
              || ||||||||||||||| ||||||||||||||||||||||  ||||||||||||||||
Db        521 AGACGCCAGCCCATCCCCAAAGATCGGCGCTCCACTGGCAAGGCCTGGGGAAAACCAGGT 580

Qy        241 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 300
                |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        581 CGCCCCTGGCCCCTATATGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 640

Qy        301 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 360
              ||||| || || ||||| |||||||||| ||||||||||||||||||||||||||||||
Db        641 CGAGGCTCTCGCCCCTCCTGGGGCCCCACTGACCCCCGGCATAGGTCGCGCAACGTGGGT 700

Qy        361 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 420
              || |||||||| |||||||||| |||||||||||||||||||||||||||||||| |||||
Db        701 AAAGTCATCGACACCCTAACGTGTGGCTTTGCCGACCTCATGGGGTACATCCCCGTCGTA 760

Qy        421 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 480
              |||||||||| | |||| |||||||||| ||||||| ||||||||||||||||||||||
Db        761 GGCGCCCCGCTTAGTGGCGCCGCCAGAGCTGTCGCGCACGGCGTGAGAGTCCTGGAGGAC 820

Qy        481 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 540
              |||||||||| |||||||||||| ||||||||| | ||||||||||||||||||||||||
Db        821 GGGGTTAATTATGCAACAGGGAACCTACCCGGTTTCCCCTTTTCTATCTTCTTGCTGGCC 880

Qy        541 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 600
              ||| ||||||||||||||   |||||||| |||||| | |||||||| | ||||| | ||
Db        881 CTGTTGTCCTGCATCACCGTTCCGGTCTCTGCTGCCCAGGTGAAGAATACCAGTAGCAGC 940

Qy        601 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 660
              |||||||||| || |||||| |||||||||||| || ||||||||| |||||| ||
Db        941 TACATGGTGACCAATGACTGCTCCAATGACAGCATCACTTGGCAGCTCGAGGCTGCGGTT 1000

Qy        661 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 720
              |||||||||||||||||||||||||||||||| ||||||||| | || | ||| ||| |
Db       1001 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAGAGTGGGGAATACGTCACGGTGTTGGGTG 1060

Qy        721 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 780
              || ||||| || ||  |||| |||| || |||||| |||||||||||||  |||||||
Db       1061 CCAGTCTCGCCAAACATGGCTGTGCGGCAGCCCGGTGCCCTCACGCAGGGTCTGCGGACG 1120

Qy        781 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 840
              ||||||| |||||||||||||||||||  |||||||||| ||||||||||||||||||||
Db       1121 CACATCGATATGGTTGTGATGTCCGCCACCTTCTGCTCTGCTCTCTACGTGGGGGACCTC 1180

Qy        841 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 900
              || || |||||||||||||| |||||  |||||| |||||||||||  ||||||||||
Db       1181 TGTGGCGGGGTGATGCTCGCGGCCCAGGTGTTCATCGTCTCGCCGCAGTACCACTGGTTT 1240

Qy        901 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 960
              || ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1241 GTGCAAGAATGCAATTGCTCCATCTACCCTGGCACCATCACTGGACACCGCATGGCATGG 1300

Qy        961 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1020
              |||||||||||||||||||||||||||||||| |||||||| |||||||| |||||| |||
Db       1301 GACATGATGATGAACTGGTCGCCCACGGCCACCATGATCCTGGCGTACGTGATGCGCGTC 1360

Qy       1021 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1080
              |||||||||| ||||||||| |||||||||||| |||||||||||||||||||||||||
Db       1361 CCCGAGGTCATCATAGACATCGTTAGCGGGGCTCACTGGGGCGTCATGTTCGGCTTGGCC 1420

Qy       1081 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1140
              |||||||||||||||||||||||||||||| ||| |||||||||||||| ||||||| |||
Db       1421 TACTTCTCTATGCAGGGAGCGTGGGCGAAGGTCATTGTCATCCTTCTGCTGGCCGCTGGG 1480

Qy       1141 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1200
              ||||||||| |||||    || ||||| || ||| || |  |||||         | ||
Db       1481 GTGGACGCGGGCACCACCACCGTTGGAGGCGCTGTTGCACGTTCCACCAACGTGATTGCC 1540

Qy       1201 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1260
              || | ||   |  |||||  |||||| || |||||| |||||||||| ||||| |||
Db       1541 GGCGTGTTCAGCCATGGCCCTCAGCAGAACATTCAGCTCATTAACACCAACGGCAGTTGG 1600

Qy       1261 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1320
              |||||||||| || ||| |||| |||||||||||||| |||||||||| ||||| |
Db       1601 CACATCAACCGTACTGCCTTGAATTGCAATGACTCCTTGAACACCGGCTTTCTCGCGGCC 1660

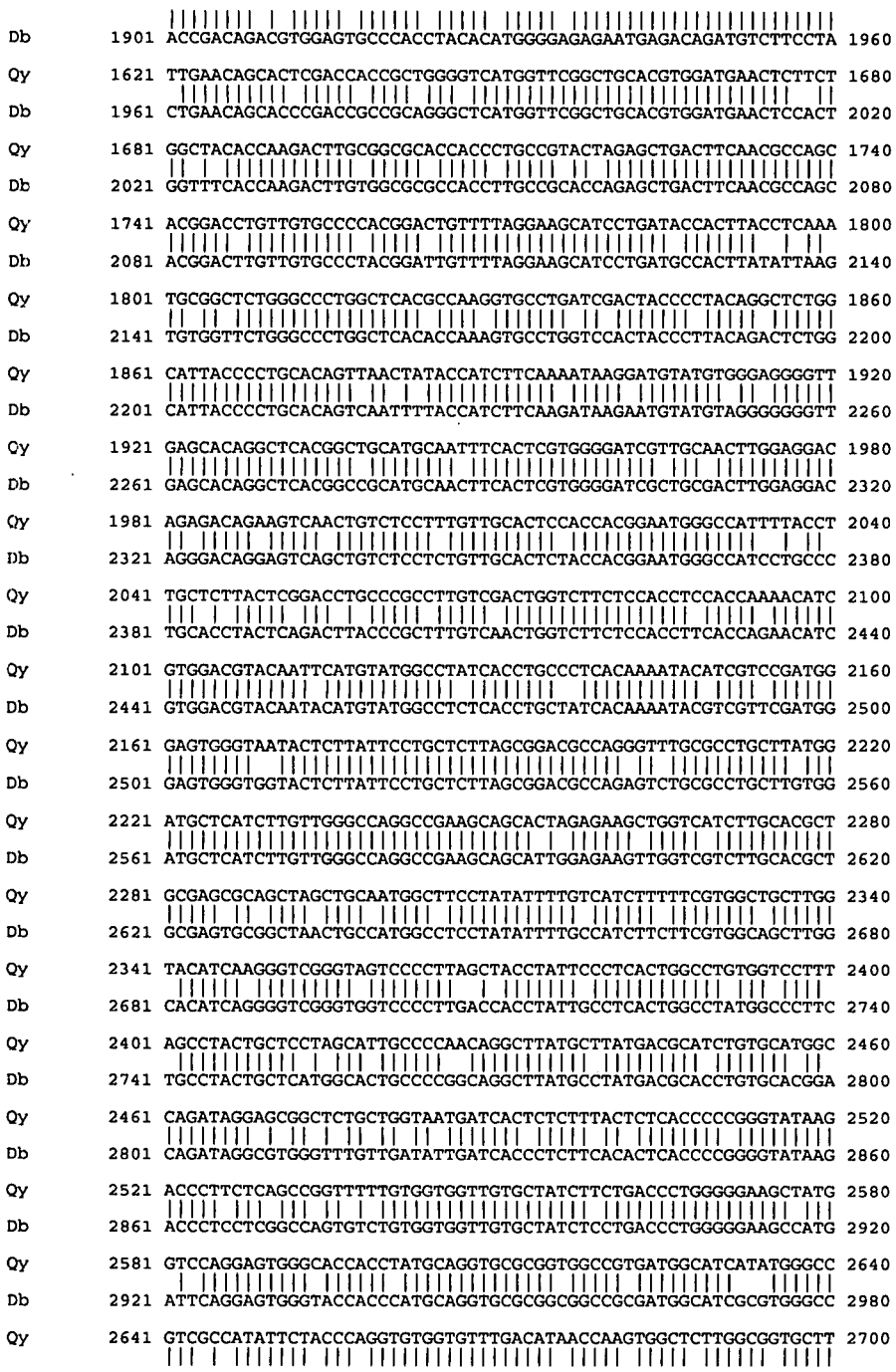
Qy       1321 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1380
              |||||||||| || |||| |||||||||| ||||| |  ||| ||||||||||||||
Db       1661 TTGTTCTACACCAACCGCTTTAACTCGTCAGGGTGTCCAGGGCGCCTGTCCGCCTGCCGC 1720

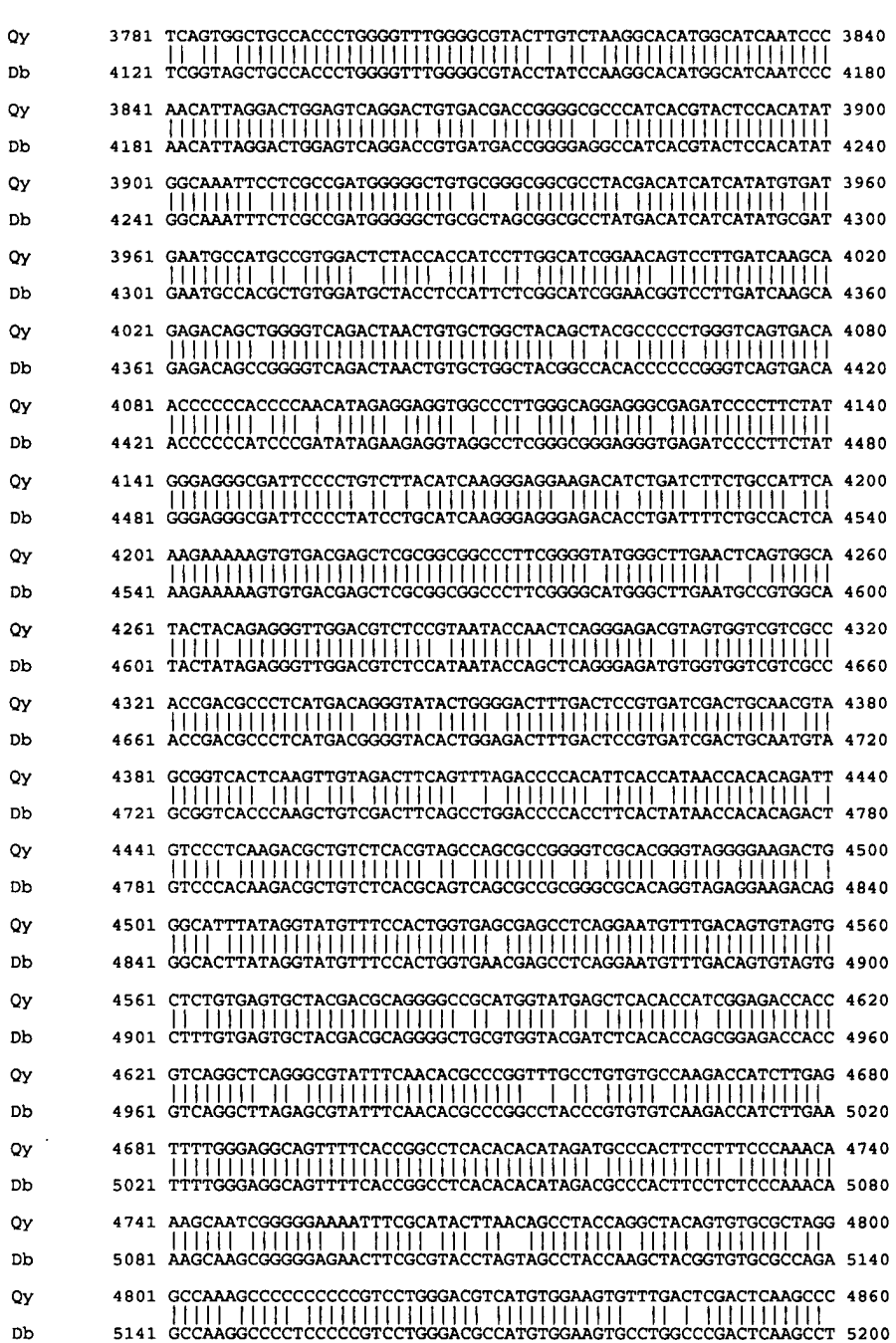
Qy       1381 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1440
              |  |||||||| |||||| | || |||||| || | || || ||||||||||||||||||
Db       1721 AACATCGAGGCTTTCCGGATAGGGTGGGGCACCCTACAGTACGAGGATAATGTCACCAAT 1780

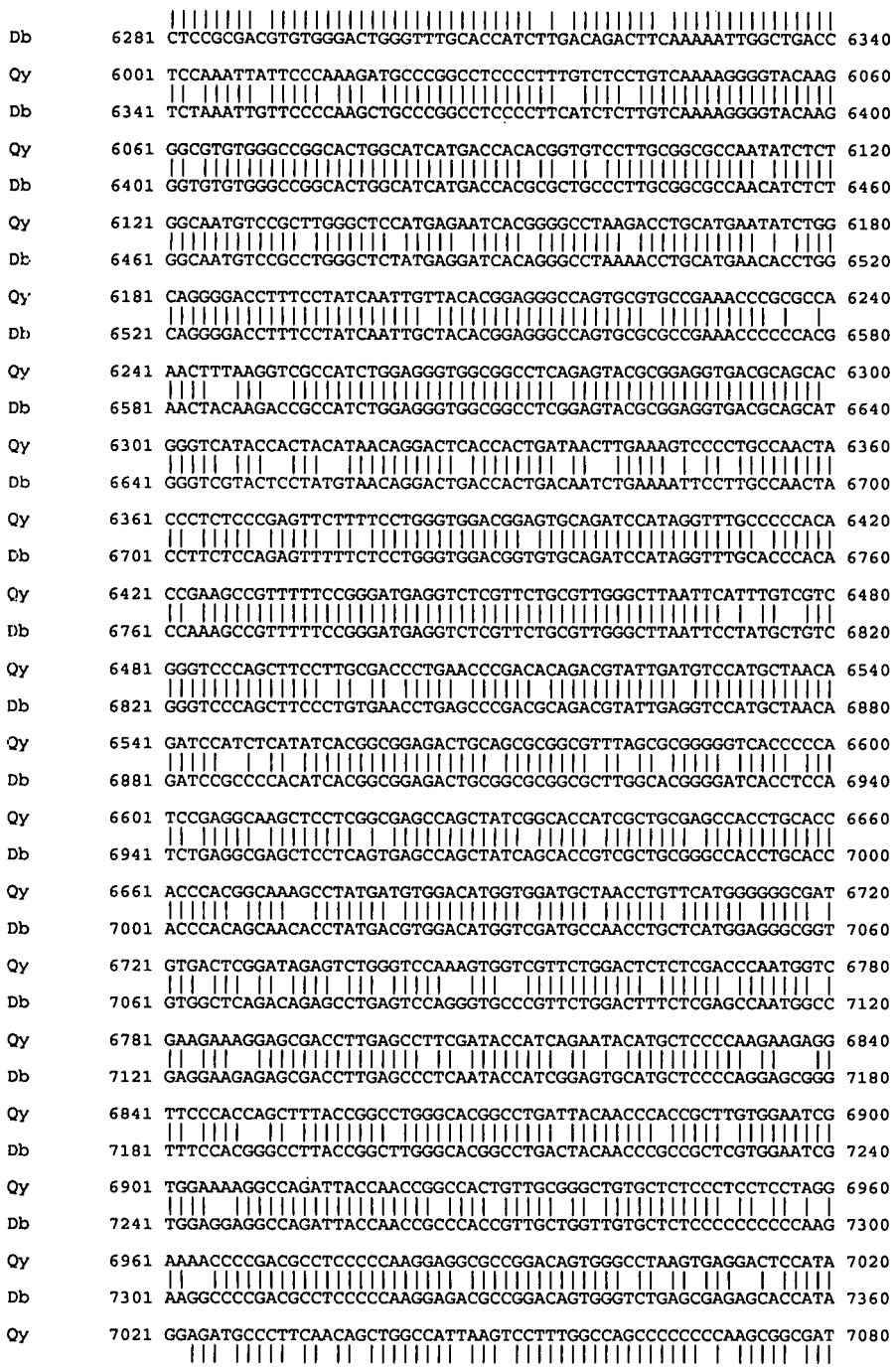
Qy       1441 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1500
              |||||||||||||| || || |||||||||||||| |||| || ||||||||| ||| ||
Db       1781 CCAGAGGATATGAGGCCGTACTGCTGGCACTACCCCCCAAAGCCGTGTGGCGTAGTCCCC 1840

Qy       1501 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1560
              |||| | ||||||||||||||||||||||||||||||||||||||| || ||||||||||
Db       1841 GCGAGGTCTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCGGTAGTAGTGGGCACG 1900

Qy       1561 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1620
```


Db 1901 ACCGACAGACGTGGAGTGCCCACCTACACATGGGGAGAGAATGAGACAGATGTCTTCCTA 1960
Qy 1621 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 1680
Db 1961 CTGAACAGCACCCGACCGCCGCAGGGCTCATGGTTCGGCTGCACGTGGATGAACTCCACT 2020
Qy 1681 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 1740
Db 2021 GGTTTCACCAAGACTTGTGGCGCGCCACCTTGCCGCACCAGAGCTGACTTCAACGCCAGC 2080
Qy 1741 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 1800
Db 2081 ACGGACTTGTTGTGCCCTACGGATTGTTTTAGGAAGCATCCTGATGCCACTTATATTAAG 2140
Qy 1801 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 1860
Db 2141 TGTGGTTCTGGGCCCTGGCTCACACCAAAGTGCCTGGTCCACTACCCTTACAGACTCTGG 2200
Qy 1861 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 1920
Db 2201 CATTACCCCTGCACAGTCAATTTTACCATCTTCAAGATAAGAATGTATGTAGGGGGGGTT 2260
Qy 1921 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 1980
Db 2261 GAGCACAGGCTCACGGCCGCATGCAACTTCACTCGTGGGGATCGCTGCGACTTGGAGGAC 2320
Qy 1981 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2040
Db 2321 AGGGACAGGAGTCAGCTGTCTCCTCTGTTGCACTCTACCACGGAATGGGCCATCCTGCCC 2380
Qy 2041 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2100
Db 2381 TGCACCTACTCAGACTTACCCGCTTTGTCAACTGGTCTTCTCCACCTTCACCAGAACATC 2440
Qy 2101 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2160
Db 2441 GTGGACGTACAATACATGTATGGCCTCTCACCTGCTATCACAAAATACGTCGTTCGATGG 2500
Qy 2161 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2220
Db 2501 GAGTGGGTGGTACTCTTATTCCTGCTCTTAGCGGACGCCAGAGTCTGCGCCTGCTTGTGG 2560
Qy 2221 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2280
Db 2561 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCATTGGAGAAGTTGGTCGTCTTGCACGCT 2620
Qy 2281 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATATTTTGTCATCTTTTTCGTGGCTGCTTGG 2340
Db 2621 GCGAGTGCGGCTAACTGCCATGGCCTCCTATATTTTGCCATCTTCTTCGTGGCAGCTTGG 2680
Qy 2341 TACATCAAGGGTCGGGTAGTCCCCTTAGCTACCTATTCCCTCACTGGCCTGTGGTCCTTT 2400
Db 2681 CACATCAGGGGTCGGGTGGTCCCCTTGACCACCTATTGCCTCACTGGCCTATGGCCCTTC 2740
Qy 2401 AGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTGCATGGC 2460
Db 2741 TGCCTACTGCTCATGGCACTGCCCCGGCAGGCTTATGCCTATGACGCACCTGTGCACGGA 2800
Qy 2461 CAGATAGGAGCGGCTCTGCTGGTAATGATCACTCTCTTTACTCTCACCCCCGGGTATAAG 2520
Db 2801 CAGATAGGCGTGGGTTTGTTGATATTGATCACCCTCTTCACACTCACCCCGGGGTATAAG 2860
Qy 2521 ACCCTTCTCAGCCGGTTTTTGTGGTGGTTGTGCTATCTTCTGACCCTGGGGGAAGCTATG 2580
Db 2861 ACCCTCCTCGGCCAGTGTCTGTGGTGGTTGTGCTATCTCCTGACCCTGGGGGAAGCCATG 2920
Qy 2581 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGCATCATATGGGCC 2640
Db 2921 ATTCAGGAGTGGGTACCACCCATGCAGGTGCGCGGCGGCCGCGATGGCATCGCGTGGGCC 2980
Qy 2641 GTCGCCATATTCTACCCAGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 2700
Db 2981 GTCACTATATTCTGCCCGGGTGTGGTGTTTGACATTACCAAATGGCTTTTGGCGTTGCTT 3040
Qy 2701 GGGCCTGCTTACCTCCTAAAAGGTGCTTTGACGCGCGTGCCGTACTTCGTCAGGGCTCAC 2760
Db 3041 GGGCCTGCTTACCTCTTAAGGGCCGCTTTGACACATGTGCCGTACTTCGTCAGAGCTCAC 3100
Qy 2761 GCTCTACTGAGGATGTGCACCATGGCAAGGCATCTCGCGGGGGGCAGGTACGTCCAGATG 2820
Db 3101 GCTCTGATAAGGGTATGCGCTTTGGTGAAGCAGCTCGCGGGGGGTAGGTATGTTCAGGTG 3160
Qy 2821 GCGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 2880
Db 3161 GCGCTATTGGCCCTTGGCAGGTGGACTGGCACCTACATCTATGACCACCTCACACCTATG 3220
Qy 2881 TCGGATTGGGCTGCTAGTGGCCTGCGGGACCTGGCGGTCGCCGTTGAGCCTATCATCTTC 2940
Db 3221 TCGGACTGGGCCGCTAGCGGCCTGCGCGACTTAGCGGTCGCCGTGGAACCCATCATCTTC 3280
Qy 2941 AGTCCGATGGAGAAGAAAGTCATTGTCTGGGGAGCGGAGACAGCTGCTTGTGGGGACATT 3000
Db 3281 AGTCCGATGGAGAAGAAGGTCATCGTCTGGGGAGCGGAGACGGCTGCATGTGGGGACATT 3340
Qy 3001 TTACACGGACTTCCCGTGTCCGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3060
Db 3341 CTACATGGACTTCCCGTGTCCGCCCGACTCGGCCAGGAGATCCTCCTCGGCCCAGCTGAT 3400
Qy 3061 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3120
Db 3401 GGCTACACCTCCAAGGGGTGGAAGCTCCTTGCTCCCATCACTGCTTATGCCCAGCAAACA 3460
Qy 3121 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCT 3180
Db 3461 CGAGGCCTCCTGGGCGCCATAGTGGTGAGTATGACGGGGCGTGACAGGACAGAACAGGCC 3520
Qy 3181 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3240
Db 3521 GGGGAAGTCCAAATCCTGTCCACAGTCTCTCAGTCCTTCCTCGGAACAACCATCTCGGGG 3580
Qy 3241 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3300
Db 3581 GTTTTGTGGACTGTTTACCACGGAGCTGGCAACAAGACTCTAGCCGGCTTACGGGGTCCG 3640
Qy 3301 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGG 3360
Db 3641 GTCACGCAGATGTACTCGAGTGCTGAGGGGGACTTGGTAGGCTGGCCCAGCCCCCCTGGG 3700
Qy 3361 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3420
Db 3701 ACCAAGTCTTTGGAGCCGTGCAAGTGTGGAGCCGTCGACCTATATCTGGTCACGCGGAAC 3760
Qy 3421 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3480
Db 3761 GCTGATGTCATCCCGGCTCGGAGACGCGGGGACAAGCGGGGAGCATTGCTCTCCCCGAGA 3820
Qy 3481 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3540
Db 3821 CCCATTTCGACCTTGAAGGGGTCCTCGGGGGGGCCGGTGCTCTGCCCTAGGGGCCACGTC 3880
Qy 3541 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3600
Db 3881 GTTGGGCTCTTCCGAGCAGCTGTGTGCTCTCGGGGCGTGGCCAAATCCATCGATTTCATC 3940
Qy 3601 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 3660
Db 3941 CCCGTTGAGACACTCGACGTTGTTACAAGGTCTCCCACTTTCAGTGACAACAGCACGCCA 4000
Qy 3661 CCTGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 3720
Db 4001 CCGGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCTCCAACTGGCAGTGGAAAG 4060
Qy 3721 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 3780
Db 4061 AGCACCAAGGTCCCTGTCGCGTATGCCGCCCAGGGGTACAAAGTACTAGTGCTTAACCCC 4120


Qy 3781 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 3840
Db 4121 TCGGTAGCTGCCACCCTGGGGTTTGGGGCGTACCTATCCAAGGCACATGGCATCAATCCC 4180
Qy 3841 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 3900
Db 4181 AACATTAGGACTGGAGTCAGGACCGTGATGACCGGGGAGGCCATCACGTACTCCACATAT 4240
Qy 3901 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 3960
Db 4241 GGCAAATTTCTCGCCGATGGGGGCTGCGCTAGCGGCGCCTATGACATCATCATATGCGAT 4300
Qy 3961 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4020
Db 4301 GAATGCCACGCTGTGGATGCTACCTCCATTCTCGGCATCGGAACGGTCCTTGATCAAGCA 4360
Qy 4021 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4080
Db 4361 GAGACAGCCGGGGTCAGACTAACTGTGCTGGCTACGGCCACACCCCCCGGGTCAGTGACA 4420
Qy 4081 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4140
Db 4421 ACCCCCCATCCCGATATAGAAGAGGTAGGCCTCGGGCGGGAGGGTGAGATCCCCTTCTAT 4480
Qy 4141 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4200
Db 4481 GGGAGGGCGATTCCCCTATCCTGCATCAAGGGAGGGAGACACCTGATTTTCTGCCACTCA 4540
Qy 4201 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4260
Db 4541 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGCATGGGCTTGAATGCCGTGGCA 4600
Qy 4261 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4320
Db 4601 TACTATAGAGGGTTGGACGTCTCCATAATACCAGCTCAGGGAGATGTGGTGGTCGTCGCC 4660
Qy 4321 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4380
Db 4661 ACCGACGCCCTCATGACGGGGTACACTGGAGACTTTGACTCCGTGATCGACTGCAATGTA 4720
Qy 4381 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4440
Db 4721 GCGGTCACCCAAGCTGTCGACTTCAGCCTGGACCCCACCTTCACTATAACCACACAGACT 4780
Qy 4441 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4500
Db 4781 GTCCCACAAGACGCTGTCTCACGCAGTCAGCGCCGCGGGCGCACAGGTAGAGGAAGACAG 4840
Qy 4501 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4560
Db 4841 GGCACTTATAGGTATGTTTCCACTGGTGAACGAGCCTCAGGAATGTTTGACAGTGTAGTG 4900
Qy 4561 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4620
Db 4901 CTTTGTGAGTGCTACGACGCAGGGGCTGCGTGGTACGATCTCACACCAGCGGAGACCACC 4960
Qy 4621 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 4680
Db 4961 GTCAGGCTTAGAGCGTATTTCAACACGCCCGGCCTACCCGTGTGTCAAGACCATCTTGAA 5020
Qy 4681 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 4740
Db 5021 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGACGCCCACTTCCTCTCCCAAACA 5080
Qy 4741 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 4800
Db 5081 AAGCAAGCGGGGGAGAACTTCGCGTACCTAGTAGCCTACCAAGCTACGGTGTGCGCCAGA 5140
Qy 4801 GCCAAAGCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 4860
Db 5141 GCCAAGGCCCCTCCCCCGTCCTGGGACGCCATGTGGAAGTGCCTGGCCCGACTCAAGCCT 5200
Qy 4861 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 4920
Db 5201 ACGCTTGCGGGCCCCACACCTCTCCTGTACCGTTTGGGCCCTATTACCAATGAGGTCACC 5260
Qy 4921 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 4980
Db 5261 CTCACACACCCTGGGACGAAGTACATCGCCACATGCATGCAAGCTGACCTTGAGGTCATG 5320
Qy 4981 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5040
Db 5321 ACCAGCACGTGGGTCCTAGCTGGAGGAGTCCTGGCAGCCGTCGCCGCATATTGCCTGGCG 5380
Qy 5041 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5100
Db 5381 ACTGGATGCGTTTCCATCATCGGCCGCTTGCACGTCAACCAGCGAGTCGTCGTTGCGCCG 5440
Qy 5101 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5160
Db 5441 GATAAGGAGGTCCTGTATGAGGCTTTTGATGAGATGGAGGAATGCGCCTCTAGGGCGGCT 5500
Qy 5161 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5220
Db 5501 CTCATCGAAGAGGGGCAGCGGATAGCCGAGATGTTGAAGTCCAAGATCCAAGGCTTGCTG 5560
Qy 5221 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5280
Db 5561 CAGCAGGCCTCTAAGCAGGCCCAGGACATACAACCCGCTATGCAGGCTTCATGGCCCAAA 5620
Qy 5281 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5340
Db 5621 GTGGAACAATTTTGGGCCAGACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5680
Qy 5341 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5400
Db 5681 GGATTGTCAACACTGCCAGGGAACCCCGCGGTGGCTTCCATGATGGCATTCAGTGCCGCC 5740
Qy 5401 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5460
Db 5741 CTCACCAGTCCGTTGTCGACCAGTACCACCATCCTTCTCAACATCATGGGAGGCTGGTTA 5800
Qy 5461 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5520
Db 5801 GCGTCCCAGATCGCACCACCCGCGGGGGCCACCGGCTTTGTCGTCAGTGGCCTGGTGGGG 5860
Qy 5521 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5580
Db 5861 GCTGCCGTGGGCAGCATAGGCCTGGGTAAGGTGCTGGTGGACATCCTGGCAGGATATGGT 5920
Qy 5581 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5640
Db 5921 GCGGGCATTTCGGGGGCCCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCTATG 5980
Qy 5641 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 5700
Db 5981 GAAGATGTCATCAATCTACTGCCTGGGATCCTGTCTCCGGGAGCCCTGGTGGTGGGGGTC 6040
Qy 5701 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 5760
Db 6041 ATCTGCGCGGCCATTCTGCGCCGCCACGTGGGACCGGGGGAGGGCGCGGTCCAATGGATG 6100
Qy 5761 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 5820
Db 6101 AACAGGCTTATTGCCTTTGCTTCCAGAGGAAACCACGTCGCCCCTACTCACTACGTGACG 6160
Qy 5821 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 5880
Db 6161 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCTCTTACTATAACCAGCCTA 6220
Qy 5881 CTCAGAAGACTCCACAACTGGATTACTGAGGACTGCCCCATCCCATGCGGCGGCTCGTGG 5940
Db 6221 CTCAGAAGACTCCACAATTGGATAACTGAGGACTGCCCCATCCCATGCTCCGGATCCTGG 6280
Qy 5941 CTCCGCGATGTGTGGGACTGGGTTTGCACCATCCTAACAGACTTTAAAAATTGGCTGACC 6000


Db 6281 CTCCGCGACGTGTGGGACTGGGTTTGCACCATCTTGACAGACTTCAAAAATTGGCTGACC 6340
Qy 6001 TCCAAATTATTCCCAAAGATGCCCGGCCTCCCCTTTGTCTCCTGTCAAAAGGGGTACAAG 6060
Db 6341 TCTAAATTGTTCCCCAAGCTGCCCGGCCTCCCCTTCATCTCTTGTCAAAAGGGGTACAAG 6400
Qy 6061 GGCGTGTGGGCCGGCACTGGCATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6120
Db 6401 GGTGTGTGGGCCGGCACTGGCATCATGACCACGCGCTGCCCTTGCGGCGCCAACATCTCT 6460
Qy 6121 GGCAATGTCCGCTTGGGCTCCATGAGAATCACGGGGCCTAAGACCTGCATGAATATCTGG 6180
Db 6461 GGCAATGTCCGCCTGGGCTCTATGAGGATCACAGGGCCTAAAACCTGCATGAACACCTGG 6520
Qy 6181 CAGGGGACCTTTCCTATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCGCCA 6240
Db 6521 CAGGGGACCTTTCCTATCAATTGCTACACGGAGGGCCAGTGCGCGCCGAAACCCCCCACG 6580
Qy 6241 AACTTTAAGGTCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6300
Db 6581 AACTACAAGACCGCCATCTGGAGGGTGGCGGCCTCGGAGTACGCGGAGGTGACGCAGCAT 6640
Qy 6301 GGGTCATACCACTACATAACAGGACTCACCACTGATAACTTGAAAGTCCCCTGCCAACTA 6360
Db 6641 GGGTCGTACTCCTATGTAACAGGACTGACCACTGACAATCTGAAAATTCCTTGCCAACTA 6700
Qy 6361 CCCTCTCCCGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCACA 6420
Db 6701 CCTTCTCCAGAGTTTTTCTCCTGGGTGGACGGTGTGCAGATCCATAGGTTTGCACCCACA 6760
Qy 6421 CCGAAGCCGTTTTTCCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6480
Db 6761 CCAAAGCCGTTTTTCCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCCTATGCTGTC 6820
Qy 6481 GGGTCCCAGCTTCCTTGCGACCCTGAACCCGACACAGACGTATTGATGTCCATGCTAACA 6540
Db 6821 GGGTCCCAGCTTCCCTGTGAACCTGAGCCCGACGCAGACGTATTGAGGTCCATGCTAACA 6880
Qy 6541 GATCCATCTCATATCACGGCGGAGACTGCAGCGCGGCGTTTAGCGCGGGGGTCACCCCCA 6600
Db 6881 GATCCGCCCCACATCACGGCGGAGACTGCGGCGCGGCGCTTGGCACGGGGATCACCTCCA 6940
Qy 6601 TCCGAGGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 6660
Db 6941 TCTGAGGCGAGCTCCTCAGTGAGCCAGCTATCAGCACCGTCGCTGCGGGCCACCTGCACC 7000
Qy 6661 ACCCACGGCAAAGCCTATGATGTGGACATGGTGGATGCTAACCTGTTCATGGGGGGCGAT 6720
Db 7001 ACCCACAGCAACACCTATGACGTGGACATGGTCGATGCCAACCTGCTCATGGAGGGCGGT 7060
Qy 6721 GTGACTCGGATAGAGTCTGGGTCCAAAGTGGTCGTTCTGGACTCTCTCGACCCAATGGTC 6780
Db 7061 GTGGCTCAGACAGAGCCTGAGTCCAGGGTGCCCGTTCTGGACTTTCTCGAGCCAATGGCC 7120
Qy 6781 GAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCAGAATACATGCTCCCCAAGAAGAGG 6840
Db 7121 GAGGAAGAGAGCGACCTTGAGCCCTCAATACCATCGGAGTGCATGCTCCCCAGGAGCGGG 7180
Qy 6841 TTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCGCTTGTGGAATCG 6900
Db 7181 TTTCCACGGGCCTTACCGGCTTGGGCACGGCCTGACTACAACCCGCCGCTCGTGGAATCG 7240
Qy 6901 TGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTCCCTCCTCCTAGG 6960
Db 7241 TGGAGGAGGCCAGATTACCAACCGCCCACCGTTGCTGGTTGTGCTCTCCCCCCCCCCAAG 7300
Qy 6961 AAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGTGAGGACTCCATA 7020
Db 7301 AAGGCCCCGACGCCTCCCCCAAGGAGACGCCGGACAGTGGGTCTGAGCGAGAGCACCATA 7360
Qy 7021 GGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCCCCAAGCGGCGAT 7080
Db 7361 TCAGAAGCCCTCCAGCAACTGGCCATCAAGACCTTTGGCCAGCCCCCCTCGAGCGGTGAT 7420
Qy 7081 TCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACGCCTCCTGATGAG 7140
Db 7421 GCAGGCTCGTCCACGGGGGCGGGCGCCGCCGAATCCGGCGGTCCGACGTCCCCTGGTGAG 7480
Qy 7141 TTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCTCGAGGGGGAGCTTGGA 7200
Db 7481 CCGGCCCCCTCAGAGACAGGTTCCGCCTCCTCTATGCCCCCCCTCGAGGGGGAGCCTGGA 7540
Qy 7201 GATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCAGGGGGGGGTGGCA 7260
Db 7541 GATCCGGACCTGGAGTCTGATCAGGTAGAGCTTCAACCTCCCCCCCAGGGGGGGGGGTA 7600
Qy 7261 GCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGACGACTCCGTCGTG 7320
Db 7601 GCTCCCGGTTCGGGCTCGGGGTCTTGGTCTACTTGCTCCGAGGAGGACGATACCACCGTG 7660
Qy 7321 TGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGTAGTCCCGAAGAG 7380
Db 7661 TGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCCTGTAGCCCCGAAGAG 7720
Qy 7381 GAGAAGTTACCGATTAACCCCTTGAGCAACTCCCTGTTGCGATATCACAACAAGGTGTAC 7440
Db 7721 GAAAAGTTGCCAATCAACCCTTTGAGTAACTCGCTGTTGCGATACCATAACAAGGTGTAC 7780
Qy 7441 TGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTTGATAGGATGCAA 7500
Db 7781 TGTACAACATCAAAGAGCGCCTCACAGAGGGCTAAAAAGGTAACTTTTGACAGGACGCAA 7840
Qy 7501 GTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCGGCCTCCAAGGTC 7560
Db 7841 GTGCTCGACGCCCATTATGACTCAGTCTTAAAGGACATCAAGCTAGCGGCTTCCAAGGTC 7900
Qy 7561 ACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCCCATTCTGCAAGA 7620
Db 7901 AGCGCAAGGCTCCTCACCTTGGAGGAGGCGTGCCAGTTGACTCCACCCCATTCTGCAAGA 7960
Qy 7621 TCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGGGCCGTTAACCAC 7680
Db 7961 TCCAAGTATGGATTCGGGGCCAAGGAGGTCCGCAGCTTGTCCGGGAGGGCCGTTAACCAC 8020
Qy 7681 ATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATTCCCACAACCATT 7740
Db 8021 ATCAAGTCCGTGTGGAAGGACCTCCTGGAAGACCCACAAACACCAATTCCCACAACCATC 8080
Qy 7741 ATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAGAAAGCAGCTCGC 7800
Db 8081 ATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCGCCAAGGGGGGTAAGAAACCAGCTCGC 8140
Qy 7801 CTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCCCTTTATGACATT 7860
Db 8141 CTCATCGTTTACCCTGACCTCGGCGTCCGGGTCTGCGAGAAAATGGCCCTCTATGACATT 8200
Qy 7861 ACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAGTATTCCCCCGCT 7920
Db 8201 ACACAAAAGCTTCCTCAGGCGGTAATGGGAGCTTCCTATGGCTTCCAGTACTCCCCTGCC 8260
Qy 7921 CAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCTATGGGTTTTTCG 7980
Db 8261 CAACGGGTGGAGTATCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCCATGGGTTTTTCG 8320
Qy 7981 TATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGGACTGAGGAGTCC 8040
Db 8321 TATGATACCCGATGCTTCGACTCAACCGTCACTGAGAGAGACATCAGGACCGAGGAGTCC 8380
Qy 8041 ATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACACTCGCTAACTGAG 8100
Db 8381 ATATACCAGGCCTGCTCCCTGCCCGAGGAGGCCCGCACTGCCATACACTCGCTGACTGAG 8440
Qy 8101 AGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGCGGGTACAGGCGT 8160
Db 8441 AGACTTTACGTAGGAGGGCCCATGTTCAACAGCAAGGGTCAAACCTGCGGTTACAGACGT 8500

```
Qy       8161 TGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACATGCTACGTGAAA 8220
             |||||||||||||||||||| |||||||||||||| |||||||||||||||||| ||||||
Db       8501 TGCCGCGCCAGCGGGGTGCTAACCACTAGCATGGGTAACACCATCACATGCTATGTGAAA 8560

Qy       8221 GCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTGGTATGCGGCGAT 8280
             ||| ||||||| || || ||||| |||||| | ||||||||||||||||||||||||||||
Db       8561 GCCCTAGCGGCCTGCAAGGCTGCGGGGATAGTTGCGCCCACAATGCTGGTATGCGGCGAT 8620

Qy       8281 GACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGGAACCTGAGAGCC 8340
             ||| | || ||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       8621 GACCTAGTAGTCATCTCAGAAAGCCAGGGGACTGAGGAGGACGAGCGGAACCTGAGAGCC 8680

Qy       8341 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCCAGACCGGAGTAT 8400
             |||||||||| |||||||||| |||||||||||||||| |||||||||||||| |||
Db       8681 TTCACGGAGGCCATGACCAGGTACTCTGCCCCTCCTGGTGATCCCCCCAGACCGGAATAT 8740

Qy       8401 GATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTGGGCCCACAAGGC 8460
             || |||||||| |||||||| || ||||||||||||||||||| ||||||| |   |||
Db       8741 GACCTGGAGCTAATAACATCCTGTTCCTCAAATGTGTCTGTGGCGTTGGGCCCGCGGGGC 8800

Qy       8461 CGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGGGCTGCCTGGGAA 8520
             |||||||||||||||||||||||||||||| ||||||||| ||||||||||||||||||||
Db       8801 CGCCGCAGATACTACCTGACCAGAGACCCAACCACTCCACTCGCCCGGGCTGCCTGGGAA 8860

Qy       8521 ACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAGTACGCCCCGACC 8580
             |||||||||||||||||| ||||||||||||||||||||||||||||||| || || |||
Db       8861 ACAGTTAGACACTCCCCTATCAATTCATGGCTGGGAAACATCATCCAGTATGCTCCAACC 8920

Qy       8581 ATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATGGCTCAAGACACG 8640
             ||||||| |||||||||||| |||||||||||||||||||||||  ||||||||
Db       8921 ATATGGGTTCGCATGGTCCTAATGACACACTTCTTCTCCATTCTCATGGTCCAAGACACC 8980

Qy       8641 CTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTGAGTCCCTTGGAC 8700
             |||||||||||||||||||||||||||| ||| | || |||||||||| ||| ||||||
Db       8981 CTGGACCAGAACCTCAACTTTGAGATGTATGGATCAGTATACTCCGTGAATCCTTTGGAC 9040

Qy       8701 CTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTGCACACATACACT 8760
             || ||||| |||||||| |||||||| |||||||||| |||||| |||||||||| ||
Db       9041 CTTCCAGCCATAATTGAGAGGTTACACGGGCTTGACGCCTTTTCTATGCACACATACTCT 9100

Qy       8761 CCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGA 8820
             | |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       9101 CACCACGAACTGACGCGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGG 9160

Qy       8821 GCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGTGGGGGGAGAGCG 8880
             | |||||||||| ||||| || |||||||||||||||||||||||||||| |||| ||||
Db       9161 GTGTGGAAGAGTCGGGCTCGCGCAGTCAGGGCGTCCCTCATCTCCCGTGGAGGGAAAGCG 9220

Qy       8881 GCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 8940
             |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db       9221 GCCGTTTGCGGCCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 9280

Qy       8941 TTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGCGCCGGCGGGGGC 9000
             |||||||| |||||||| ||||| |||||||| ||||||||||||||||||||||||||||
Db       9281 TTGCCGGAGGCGCGCCTACTGGACTTATCCAGTTGGTTCACCGTCGGCGCCGGCGGGGGC 9340

Qy       9001 GACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTTGGCCTACTCCTA 9060
             ||||||| |||||||||||||| |||||||||||| ||| ||||| ||||||||||||
Db       9341 GACATTTTTCACAGCGTGTCGCGCGCCCGACCCCGCTCATTACTCTTCGGCCTACTCCTA 9400

Qy       9061 CTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCGG 9099
             ||||| |||||||||||||| ||||||||||||||||||
Db       9401 CTTTTCGTAGGGGTAGGCCTCTTCCTACTCCCCGCTCGG 9439


RESULT 6
AAQ38221
ID   AAQ38221 standard; cDNA to mRNA; 9511 BP.
XX
AC   AAQ38221;
XX
DT   27-AUG-2003  (revised)
DT   25-MAR-2003  (revised)
DT   01-JUL-1993  (first entry)
XX
DE   NANBH virus strain HC-J8 genome cDNA sequence.
XX
KW   Non A non B hepatitis virus; plasma; degenerate; ss.
XX
OS   Non-A.
OS   non-B hepatitis virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             342. .9443
FT                   /*tag= a
XX
PN   EP532167-A2.
XX
PD   17-MAR-1993.
XX
PF   30-JUL-1992;   92EP-00306952.
XX
PR   09-AUG-1991;   91JP-00287402.
PR   05-DEC-1991;   91JP-00360441.
XX
PA   (IMMO ) IMMUNO JAPAN INC.
XX
PI   Okamoto H,  Nakamura T;
XX
DR   WPI; 1993-087166/11.
DR   N-PSDB; AAR33539, AAR33214.
XX
PT   Polynucleotide(s), polypeptide(s) and antibodies of NANBH virus - useful
PT   for detecting NANBH, as a vaccine and for screening blood samples.
XX
PS   Claim 7; Page 59-64; 93pp; English.
XX
CC   RNA was isolated from the plasma of human patients positive for NANBH
CC   virus (strain HC-J8) and was subjected to reverse transcription to
CC   produce cDNA. The resulting cDNA was amplified by PCR, and nucleic acid
CC   sequences determined by analysis of clones obtd. by PCR amplification (42
CC   clones in total). The NANBH HC-J8 degenerate genome was found to contain
CC   a degenerate open reading frame encoding polypeptide precursors of 3033
CC   amino acid residues. See also AAQ38172-220. (Updated on 25-MAR-2003 to
CC   correct PN field.) (Updated on 27-AUG-2003 to correct OS field.)
XX
SQ   Sequence 9511 BP; 2020 A; 2709 C; 2539 G; 2108 T; 0 U; 135 Other;

  Query Match             62.1%;  Score 5652.2;  DB 2;  Length 9511;
  Best Local Similarity   75.5%;  Pred. No. 0;
  Matches 6867;  Conservative  104;  Mismatches 2128;  Indels    0;  Gaps    0;

Qy          1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 60
              |||||||||||||||||||||||||||||||||||||||||||||| ||||| ||||||||
Db        342 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACAAACCGCCGCCCACAG 401

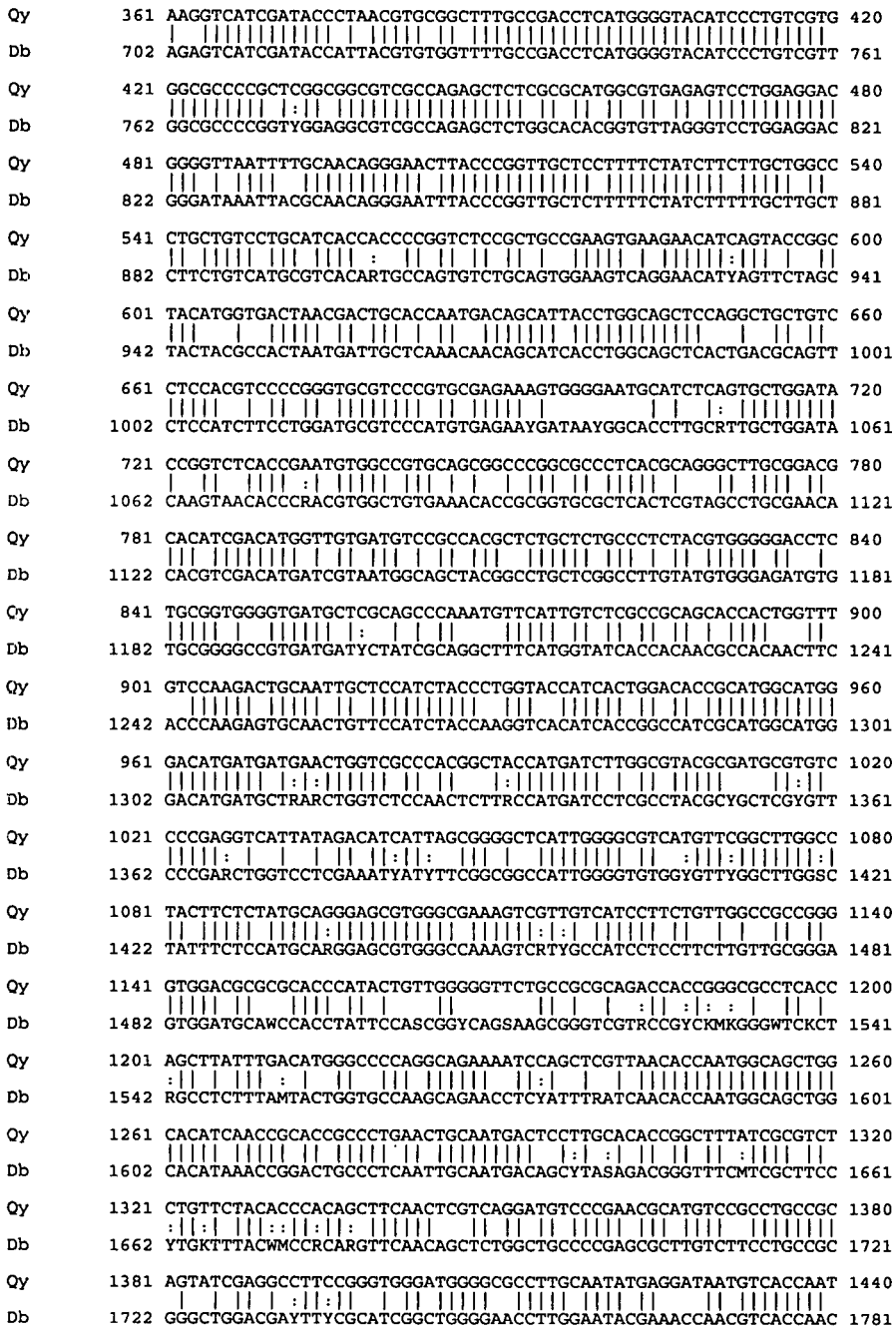
Qy         61 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 120
              |||||||||| ||||| ||||| |||||||||||||||| |||||| ||||||||||
Db        402 GACGTTAAGTTCCCGGGTGGCGGTCAGATCGTTGGCGGAGTTTACTTGCTGCCGCGCAGG 461

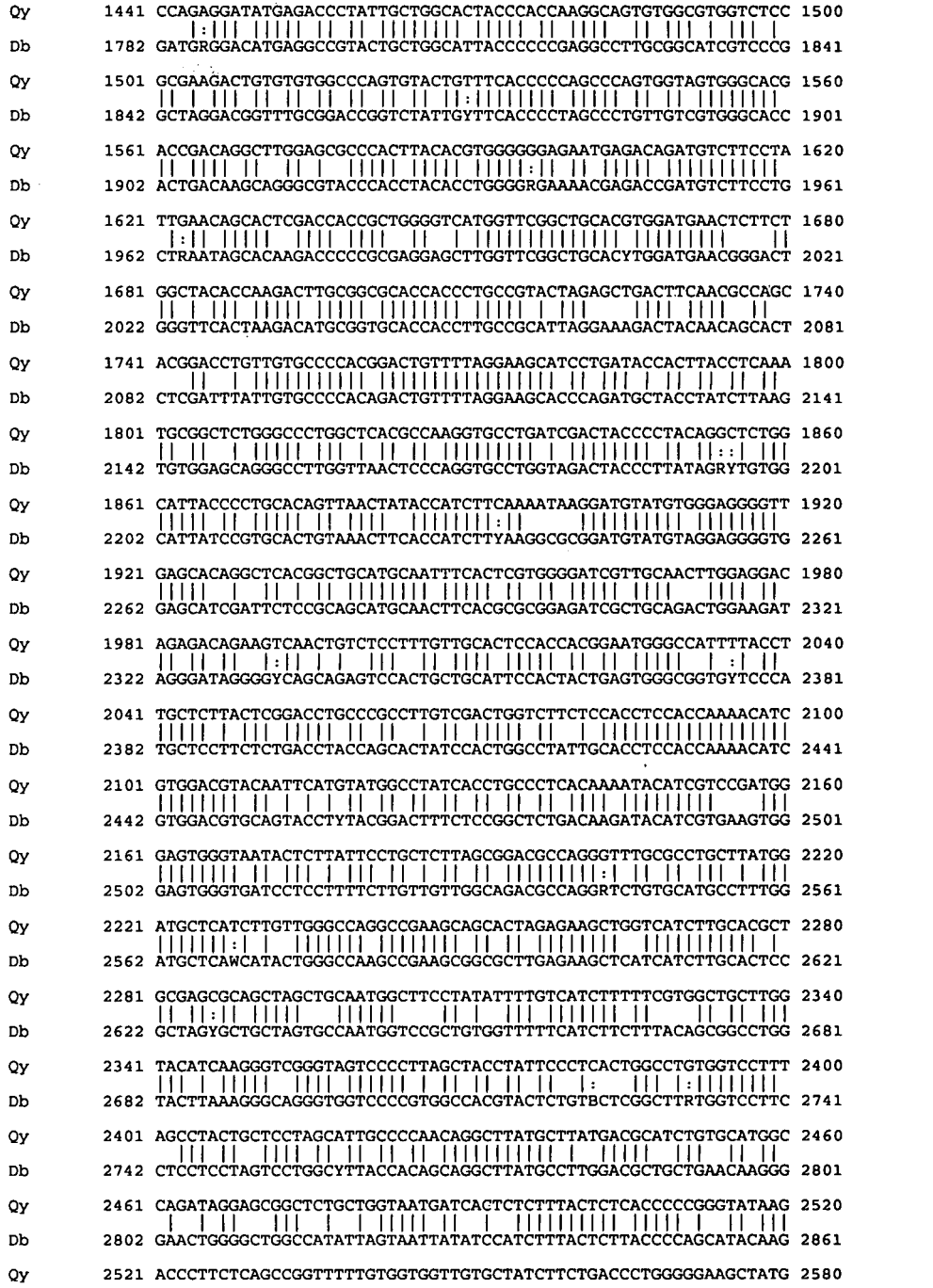
Qy        121 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 180
              |||||||||||||||||||||||||||||||||||||||| ||||| |||||||| ||||||
Db        462 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCYGAGCGATCCCAGCCGCGTGGA 521

Qy        181 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 240
              | |||||||||||||| |||||||||||||||| ||||| |||||||||| ||||||
Db        522 CGACGCCAGCCCATCCCGAAAGATCGGCGCTCCACCGGCAAGTCCTGGGGAAAGCCAGGA 581

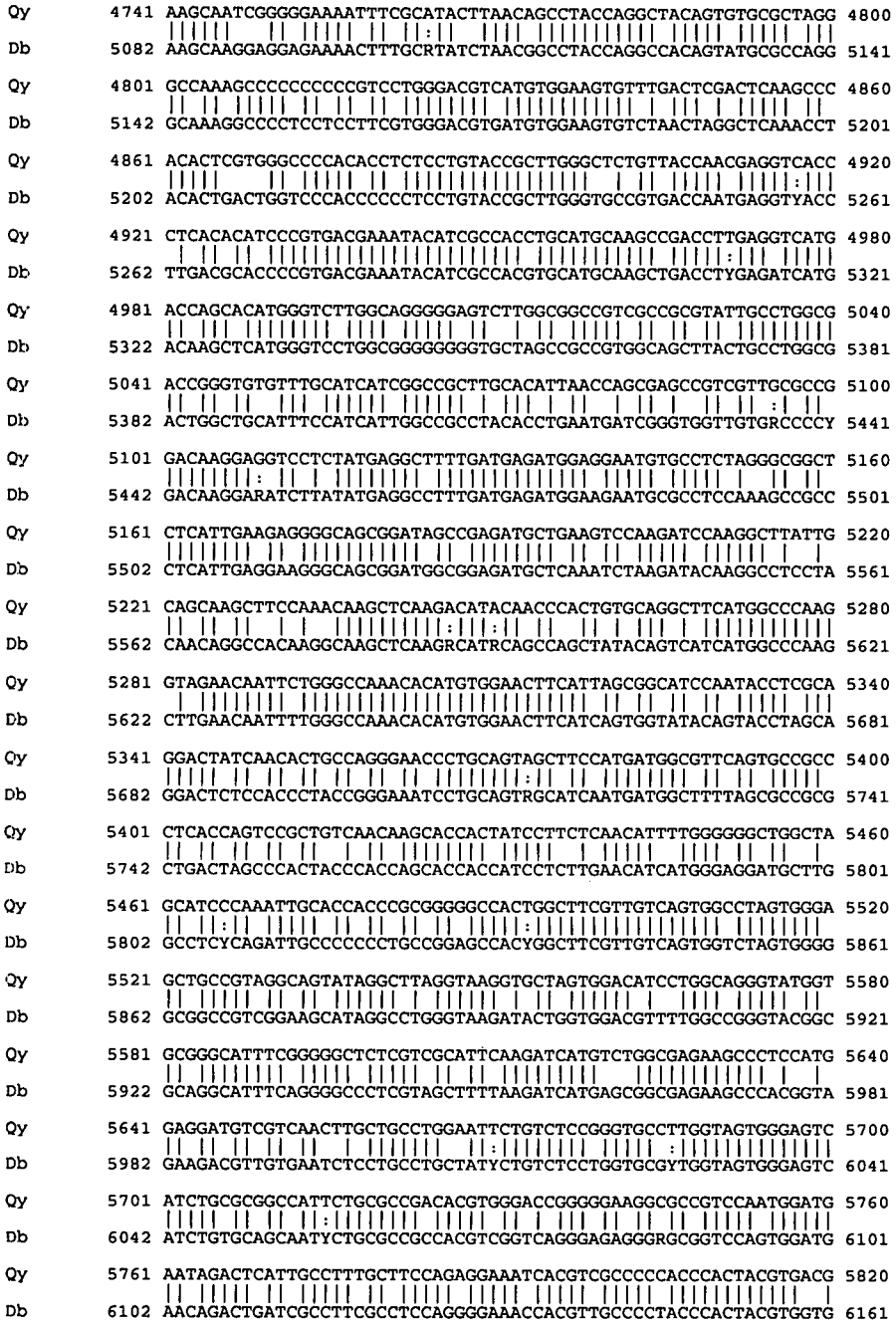
Qy        241 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 300
              || || |||||||| ||||| || |||||   ||||||||| || ||||||||||||||
Db        582 TATCCTTGGCCCCTGTACGGAAACGAGGGTTGCGGCTGGGCGGGTTGGCTCCTGTCCCCC 641

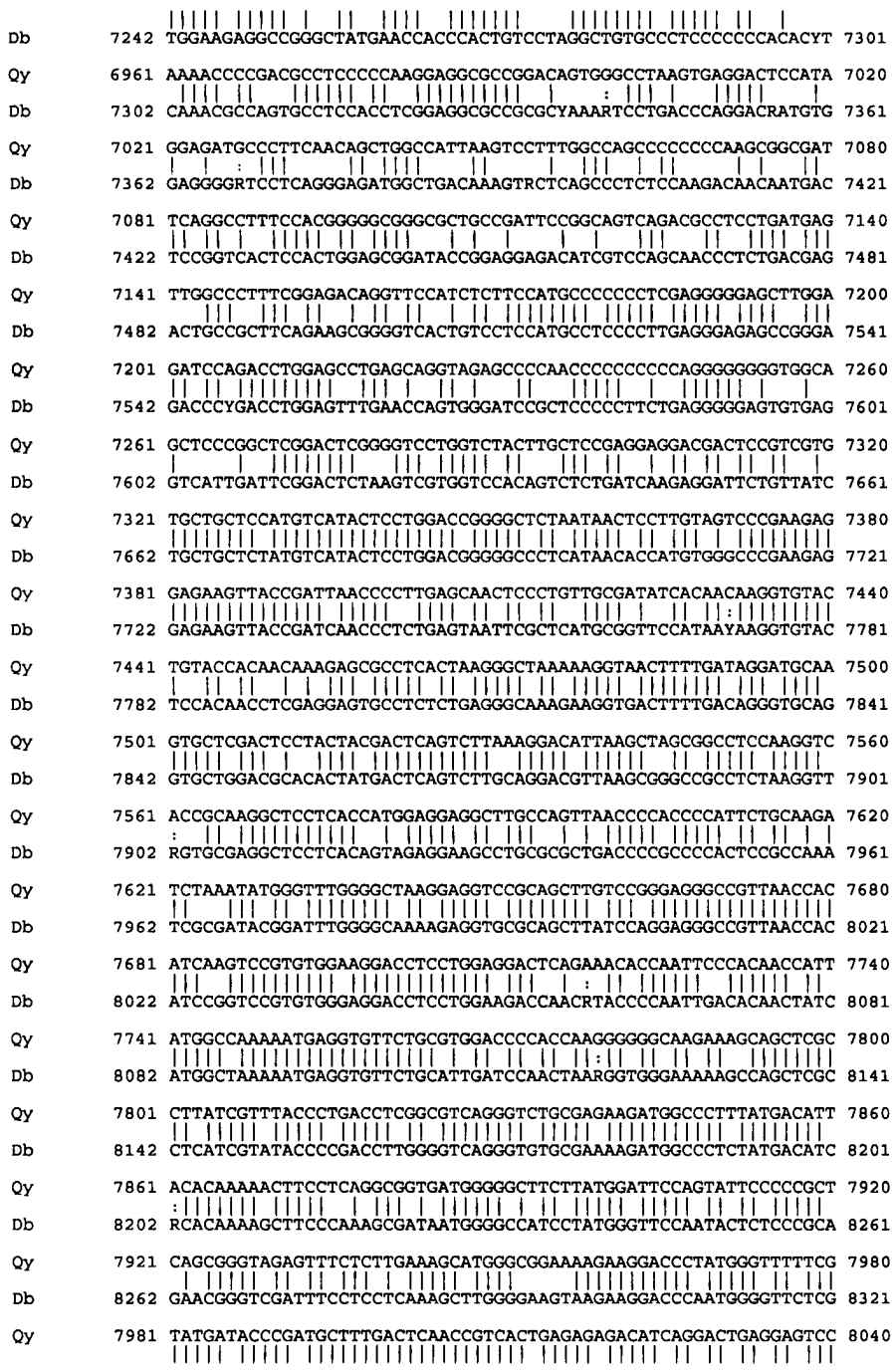
Qy        301 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 360
              || || || |||||  |||||||||||  |||||||||||||| || |||||  ||||
Db        642 CGCGGGTCTCGTCCTACTTGGGGCCCCACCGACCCCCGGCATAGATCACGCAATTTGGGC 701
```








Qy 4741 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 4800
Db 5082 AAGCAAGGAGGAGAAAACTTTGCRTATCTAACGGCCTACCAGGCCACAGTATGCGCCAGG 5141
Qy 4801 GCCAAAGCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 4860
Db 5142 GCAAAGGCCCCTCCTCCTTCGTGGGACGTGATGTGGAAGTGTCTAACTAGGCTCAAACCT 5201
Qy 4861 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 4920
Db 5202 ACACTGACTGGTCCCACCCCCCTCCTGTACCGCTTGGGTGCCGTGACCAATGAGGTYACC 5261
Qy 4921 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 4980
Db 5262 TTGACGCACCCCGTGACGAAATACATCGCCACGTGCATGCAAGCTGACCTYGAGATCATG 5321
Qy 4981 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5040
Db 5322 ACAAGCTCATGGGTCCTGGCGGGGGGGGTGCTAGCCGCCGTGGCAGCTTACTGCCTGGCG 5381
Qy 5041 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5100
Db 5382 ACTGGCTGCATTTCCATCATTGGCCGCCTACACCTGAATGATCGGGTGGTTGTGRCCCCY 5441
Qy 5101 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5160
Db 5442 GACAAGGARATCTTATATGAGGCCTTTGATGAGATGGAAGAATGCGCCTCCAAAGCCGCC 5501
Qy 5161 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5220
Db 5502 CTCATTGAGGAAGGGCAGCGGATGGCGGAGATGCTCAAATCTAAGATACAAGGCCTCCTA 5561
Qy 5221 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5280
Db 5562 CAACAGGCCACAAGGCAAGCTCAAGRCATRCAGCCAGCTATACAGTCATCATGGCCCAAG 5621
Qy 5281 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5340
Db 5622 CTTGAACAATTTTGGGCCAAACACATGTGGAACTTCATCAGTGGTATACAGTACCTAGCA 5681
Qy 5341 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5400
Db 5682 GGACTCTCCACCCTACCGGGAAATCCTGCAGTRGCATCAATGATGGCTTTTAGCGCCGCG 5741
Qy 5401 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5460
Db 5742 CTGACTAGCCCACTACCCACCAGCACCACCATCCTCTTGAACATCATGGGAGGATGCTTG 5801
Qy 5461 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5520
Db 5802 GCCTCYCAGATTGCCCCCCCTGCCGGAGCCACYGGCTTCGTTGTCAGTGGTCTAGTGGGG 5861
Qy 5521 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5580
Db 5862 GCGGCCGTCGGAAGCATAGGCCTGGGTAAGATACTGGTGGACGTTTTGGCCGGGTACGGC 5921
Qy 5581 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5640
Db 5922 GCAGGCATTTCAGGGGCCCTCGTAGCTTTTAAGATCATGAGCGGCGAGAAGCCCACGGTA 5981
Qy 5641 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 5700
Db 5982 GAAGACGTTGTGAATCTCCTGCCTGCTATYCTGTCTCCTGGTGCGYTGGTAGTGGGAGTC 6041
Qy 5701 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 5760
Db 6042 ATCTGTGCAGCAATYCTGCGCCGCCACGTCGGTCAGGGAGAGGGRGCGGTCCAGTGGATG 6101
Qy 5761 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 5820
Db 6102 AACAGACTGATCGCCTTCGCCTCCAGGGGAAACCACGTTGCCCCTACCCACTACGTGGTG 6161
Qy 5821 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 5880
Db 6162 GAGTCTGACGCTTCACAGCGTGTRACGCAGGTGCTGAGTTCACTTACAATTACCAGCTTA 6221
Qy 5881 CTCAGAAGACTCCACAACTGGATTACTGAGGACTGCCCCATCCCATGCGGCGGCTCGTGG 5940
Db 6222 CTTAGGAGACTACATGCCTGGATCACTGAAGATTGCCCARTCCCATGCTCGGGGTCTTGG 6281
Qy 5941 CTCCGCGATGTGTGGGACTGGGTTTGCACCATCCTAACAGACTTTAAAAATTGGCTGACC 6000
Db 6282 CTCCAGGACATTTGGGATTGGGTTTGTTCCATCCTCACAGACTTYAAAAACTGGCTGTCT 6341
Qy 6001 TCCAAATTATTCCCAAAGATGCCCGGCCTCCCCTTTGTCTCCTGTCAAAAGGGGTACAAG 6060
Db 6342 TCAAAATTACTCCCCAAGATGCCCGGCATTCCCTTTATCTCTTGCCAGAAGGGATACAAG 6401
Qy 6061 GGCGTGTGGGCCGGCACTGGCATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6120
Db 6402 GGTGTATGGGCTGGTACGGGTGTCATGACYACTCGRTRCCCATGTGGAGCAAACATCTCG 6461
Qy 6121 GGCAATGTCCGCTTGGGCTCCATGAGAATCACGGGGCCTAAGACCTGCATGAATATCTGG 6180
Db 6462 GGCCATGTCCGCATGGGCACCATGAAAATAACAGGCCCGAAGACTTGCTTGAACCTGTGG 6521
Qy 6181 CAGGGGACCTTTCCTATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCGCCA 6240
Db 6522 CAGGGGACTTTCCCCATTAATTGTTACACAGAAGGGCCYTGCGTGCCAAAACCCCCTCCT 6581
Qy 6241 AACTTTAAGGTCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6300
Db 6582 AATTACAAGACCGCAATTTGGAGGGTGGCAGCGTCGGAGTACGTTGAGGTCACACAGCAT 6641
Qy 6301 GGGTCATACCACTACATAACAGGACTCACCACTGATAACTTGAAAGTCCCCTGCCAACTA 6360
Db 6642 GGCTCTTTCTCGTATGTAACRGGGTTAACCAGTGACAACCTTAAGGTYCCTTGCCAGGTA 6701
Qy 6361 CCCTCTCCCGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCACA 6420
Db 6702 CCAGCTCCAGAATTTTTCTCTTGGGTGGACGGGGTGCAAATCCACCGATTCGCCCCCGTW 6761
Qy 6421 CCGAAGCCGTTTTTCCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6480
Db 6762 CCAGGTCCCTTCTTTCGGGATGAGGTAACGTTCACCGTAGGCCTTAACTCCTTCGTGGTC 6821
Qy 6481 GGGTCCCAGCTTCCTTGCGACCCTGAACCCGACACAGACGTATTGATGTCCATGCTAACA 6540
Db 6822 GGCTCTCAGCTCCCTTGCGATCCTGAGCCGGACACCGARGTACTGGCCTCYATGTTGACA 6881
Qy 6541 GATCCATCTCATATCACGGCGGAGACTGCAGCGCGGCGTTTAGCGCGGGGGTCACCCCCA 6600
Db 6882 GACCCGTCCCACATCACCGCKGAGGCGGCAGCCAGGCGATTGGCAAGGGGATCTCCCCCY 6941
Qy 6601 TCCGAGGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 6660
Db 6942 TCACAGGCTAGCTCCTCAGCGAGCCAGCTCTCTGCCCCGTCCTTGAAGGCTACCTGTACC 7001
Qy 6661 ACCCACGGCAAAGCCTATGATGTGGACATGGTGGATGCTAACCTGTTCATGGGGGGCGAT 6720
Db 7002 ACCCATAAGACAGCATATGATTGTGACATGGTGGATGCYAACCTTTTCATGGGAGGMGAT 7061
Qy 6721 GTGACTCGGATAGAGTCTGGGTCCAAAGTGGTCGTTCTGGACTCTCTCGACCCAATGGTC 6780
Db 7062 GTGAYCCGGATTGAGTCTGACTCTAAGGTGATCGTTCTAGACTCCCTCGATTCCATGACT 7121
Qy 6781 GAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCAGAATACATGCTCCCCAAGAAGAGG 6840
Db 7122 GAGGTAGAGGATGATCGTGAGCCTTCTGTACCATCAGAGTACCTGATCAAGAGGAGAAAG 7181
Qy 6841 TTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCGCTTGTGGAATCG 6900
Db 7182 TTCCCACCGGCGCTGCCTCCTTGGGCCCGTCCAGACTACAATCCTGTTTTGATCGAGACA 7241
Qy 6901 TGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTCCCTCCTCCTAGG 6960


Db 7242 TGGAAGAGGCCGGGCTATGAACCACCCACTGTCCTAGGCTGTGCCCTCCCCCCCACACYT 7301
Qy 6961 AAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGTGAGGACTCCATA 7020
Db 7302 CAAACGCCAGTGCCTCCACCTCGGAGGCGCCGCGCYAAARTCCTGACCCAGGACRATGTG 7361
Qy 7021 GGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCCCCCAAGCGGCGAT 7080
Db 7362 GAGGGGRTCCTCAGGGAGATGGCTGACAAAGTRCTCAGCCCTCTCCAAGACAACAATGAC 7421
Qy 7081 TCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACGCCTCCTGATGAG 7140
Db 7422 TCCGGTCACTCCACTGGAGCGGATACCGGAGGAGACATCGTCCAGCAACCCTCTGACGAG 7481
Qy 7141 TTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCTCGAGGGGGAGCTTGGA 7200
Db 7482 ACTGCCGCTTCAGAAGCGGGGTCACTGTCCTCCATGCCTCCCCTTGAGGGAGAGCCGGGA 7541
Qy 7201 GATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCCAGGGGGGGGTGGCA 7260
Db 7542 GACCCYGACCTGGAGTTTGAACCAGTGGGATCCGCTCCCCCTTCTGAGGGGGAGTGTGAG 7601
Qy 7261 GCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGACGACTCCGTCGTG 7320
Db 7602 GTCATTGATTCGGACTCTAAGTCGTGGTCCACAGTCTCTGATCAAGAGGATTCTGTTATC 7661
Qy 7321 TGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGTAGTCCCGAAGAG 7380
Db 7662 TGCTGCTCTATGTCATACTCCTGGACGGGGGCCCTCATAACACCATGTGGGCCCGAAGAG 7721
Qy 7381 GAGAAGTTACCGATTAACCCCTTGAGCAACTCCCTGTTGCGATATCACAACAAGGTGTAC 7440
Db 7722 GAGAAGTTACCGATCAACCCTCTGAGTAATTCGCTCATGCGGTTCCATAAYAAGGTGTAC 7781
Qy 7441 TGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTTGATAGGATGCAA 7500
Db 7782 TCCACAACCTCGAGGAGTGCCTCTCTGAGGGCAAAGAAGGTGACTTTTGACAGGGTGCAG 7841
Qy 7501 GTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCGGCCTCCAAGGTC 7560
Db 7842 GTGCTGGACGCACACTATGACTCAGTCTTGCAGGACGTTAAGCGGGCCGCCTCTAAGGTT 7901
Qy 7561 ACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCCCATTCTGCAAGA 7620
Db 7902 RGTGCGAGGCTCCTCACAGTAGAGGAAGCCTGCGCGCTGACCCCGCCCCACTCCGCCAAA 7961
Qy 7621 TCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGGGCCGTTAACCAC 7680
Db 7962 TCGCGATACGGATTTGGGGCAAAAGAGGTGCGCAGCTTATCCAGGAGGGCCGTTAACCAC 8021
Qy 7681 ATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATTCCCACAACCATT 7740
Db 8022 ATCCGGTCCGTGTGGGAGGACCTCCTGGAAGACCAACRTACCCCAATTGACACAACTATC 8081
Qy 7741 ATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAGAAAGCAGCTCGC 7800
Db 8082 ATGGCTAAAAATGAGGTGTTCTGCATTGATCCAACTAARGGTGGGAAAAAGCCAGCTCGC 8141
Qy 7801 CTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCCCTTTATGACATT 7860
Db 8142 CTCATCGTATACCCCGACCTTGGGGTGTGCGAAAAGATGGCCCTCTATGACATC 8201
Qy 7861 ACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAGTATTCCCCCGCT 7920
Db 8202 RCACAAAAGCTTCCCAAAGCGATAATGGGGCCATCCTATGGGTTCCAATACTCTCCCGCA 8261
Qy 7921 CAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCTATGGGTTTTTCG 7980
Db 8262 GAACGGGTCGATTTCCTCCTCAAAGCTTGGGGAAGTAAGAAGGACCCAATGGGGTTCTCG 8321
Qy 7981 TATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGGACTGAGGAGTCC 8040
Db 8322 TATGACACCCGCTGCTTTGACTCAACCGTCACGGAGAGGGACATAAGAACAGAAGAATCC 8381
Qy 8041 ATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACACTCGCTAACTGAG 8100
Db 8382 ATATATCAGGCTTGTTCTCTGCCTCAAGAAGCCAGAACTGTCATACACTCGCTCACTGAG 8441
Qy 8101 AGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGCGGGTACAGGCGT 8160
Db 8442 AGACTTTACGTAGGAGGGCCCATGACAAACAGCAAAGGGCAATCCTGCGGCTACAGGCGT 8501
Qy 8161 TGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACATGCTACGTGAAA 8220
Db 8502 TGCCGCGCAAGCGGKGTTTTCACCACCAGCATGGGGAATACCATGACATGTTACATCAAA 8561
Qy 8221 GCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTGGTATGCGGCGAT 8280
Db 8562 GCCCTTGCAGCGTGTAAGGCTGCRGGGATCGTGGACCCTGTTATGTTGGTGTGTGGAGAC 8621
Qy 8281 GACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGGAACCTGAGAGCC 8340
Db 8622 GACCTGGTCGTCATCTCAGAGAGCCAAGGTAACGAGGAGGACGAGCGAAACCTGAGAGCT 8681
Qy 8341 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCCAGACCGGAGTAT 8400
Db 8682 TTCACGGAGGCTATGACCAGGTATTCCGCCCCTCCCGGTGACCTTCCCAGACCGGAATAT 8741
Qy 8401 GATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTGGGCCCACAAGGC 8460
Db 8742 GACTTGGAGCTTATAACATCCTGCTCCTCAAACGTATCGGTAGCGCTGGACTCTCGGGGT 8801
Qy 8461 CGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGGGCTGCCTGGGAA 8520
Db 8802 CGCCGCCGGTACTTCCTAACCAGAGACCCTACCACTCCAATCACCCGAGCTGCTTGGGAA 8861
Qy 8521 ACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAGTACGCCCCGACC 8580
Db 8862 ACAGTAAGACACTCCCCTGTCAATTCTTGGCTGGGCAACATCATCCAGTACGCCCCCACA 8921
Qy 8581 ATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATGGCTCAAGACACG 8640
Db 8922 ATCTGGGTCCGGATGGTCATAATGACTCACTTCTTCTCCATACTATTGGCCCAGGACACT 8981
Qy 8641 CTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTGAGTCCCTTGGAC 8700
Db 8982 CTGAACCAAAATCTCAATTTTGAGATGTACGGGGCAGTATACTCGGTCAATCCATTAGAC 9041
Qy 8701 CTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTGCACACATACACT 8760
Db 9042 CTACCGGCCATAATTGAAAGGCTACATGGGCTTGAAGCCTTTTCACTGCACACATACTCT 9101
Qy 8761 CCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGA 8820
Db 9102 CCCCACGAACTCTCACGGGTGGCAGCAACTCTCAGAAAACTTGGAGCGCCTCCCCTTAGA 9161
Qy 8821 GCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGTGGGGGGAGAGCG 8880
Db 9162 GCGTGGAAGAGTCGGGCGCGTGCCGTGAGAGCTTCACTCATCGCCCAAGGAGCGAGGGCG 9221
Qy 8881 GCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 8940
Db 9222 GCCATTTGTGGCCGCTACCTCTTCAACTGGGCGGTGAAAACAAAGCTCAAACTCACTCCA 9281
Qy 8941 TTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGCGCCGGCGGGGGC 9000
Db 9282 TTGCCCGAGGCGAGCCGCCTGGATTTATCCGGGTGGTTCACCGTGGGCGCCGGCGGGGGC 9341
Qy 9001 GACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTTGGCCTACTCCTA 9060
Db 9342 GACATTTATCACAGCGTGTCGCATGCYCGACCCCGCCTATTACTCCTTTGCCTACTCCTA 9401
Qy 9061 CTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCGG 9099
Db 9402 CTTAGCGTAGGAGTAGGCATCTTTTTACTCCCCGCTCGG 9440

```
RESULT 7
AAC86645
ID   AAC86645 standard; DNA; 9611 BP.
XX
AC   AAC86645;
XX
DT   02-APR-2001  (first entry)
XX
DE   Nucleotide sequence of chimeric Hepatitis C virus clone pH77CV-J6S.
XX
KW   HCV; HCV strain HC-J6CH; HCV genotype 2a; antiviral; vaccine; ss.
XX
OS   Synthetic.
OS   Hepatitis C virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             342. .9389
FT                   /*tag=  a
XX
PN   WO200075338-A2.
XX
PD   14-DEC-2000.
XX
PF   02-JUN-2000; 2000WO-US015446.
XX
PR   04-JUN-1999;   99US-0137693P.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Yanagi M,  Bukh J,  Emerson SU,  Purcell RH;
XX
DR   WPI; 2001-061728/07.
XX
PT   Nucleic acid molecule encoding human hepatitis C virus of genotype 2a for
PT   developing vaccines, for diagnosis of hepatitis C virus and in screening
PT   assays for identification of antiviral agents.
XX
PS   Disclosure; Page 100-103; 167pp; English.
XX
CC   AAC86645-48 represent chimeric cDNA clones of infectious Hepatitis C
CC   virus (HCV). In each clone the C, E1 and E2 genes are derived from HCV
CC   strain HC-J6CH genotype 2a. The p7 protein was derived from HCV strain HC
CC   -J6CH or HCV strain pCV-H77C, and the NS genes were derived from HCV
CC   strain pCV-H77C. Such HCV sequences are capable of expressing the virus
CC   when transfected into cells. The HCV protein is useful for assaying
CC   candidate antiviral agents for activity against HCV. Antibodies specific
CC   for HCV polypeptide are useful in prevention and treatment of diseases
CC   caused by HCV in animals, in particular humans. The HCV polypeptides
CC   serve as immunogens in the development of vaccines for preventing HCV in
CC   mammals or as antigens in diagnostic assays for detecting the presence of
CC   HCV in biological samples. The HCV polynucleotide is also useful for
CC   identifying cell lines capable of supporting the replication of HCV in
CC   vitro and to produce attenuated viral strains via passage in vitro or in
CC   vivo
XX
SQ   Sequence 9611 BP; 1935 A; 2896 C; 2682 G; 2098 T; 0 U; 0 Other;

  Query Match             59.3%;  Score 5399.2;  DB 5;  Length 9611;
  Best Local Similarity   75.0%;  Pred. No. 0;
  Matches 6829;  Conservative    0;  Mismatches 2203;  Indels   78;  Gaps    3;

Qy          1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        342 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 401

Qy         61 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        402 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 461

Qy        121 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        462 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 521

Qy        181 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        522 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 581

Qy        241 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        582 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 641

Qy        301 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        642 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 701

Qy        361 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        702 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 761

Qy        421 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        762 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 821

Qy        481 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        822 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 881

Qy        541 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        882 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 941

Qy        601 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        942 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 1001

Qy        661 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1002 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 1061

Qy        721 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1062 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 1121

Qy        781 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1122 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 1181

Qy        841 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1182 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 1241

Qy        901 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1242 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 1301

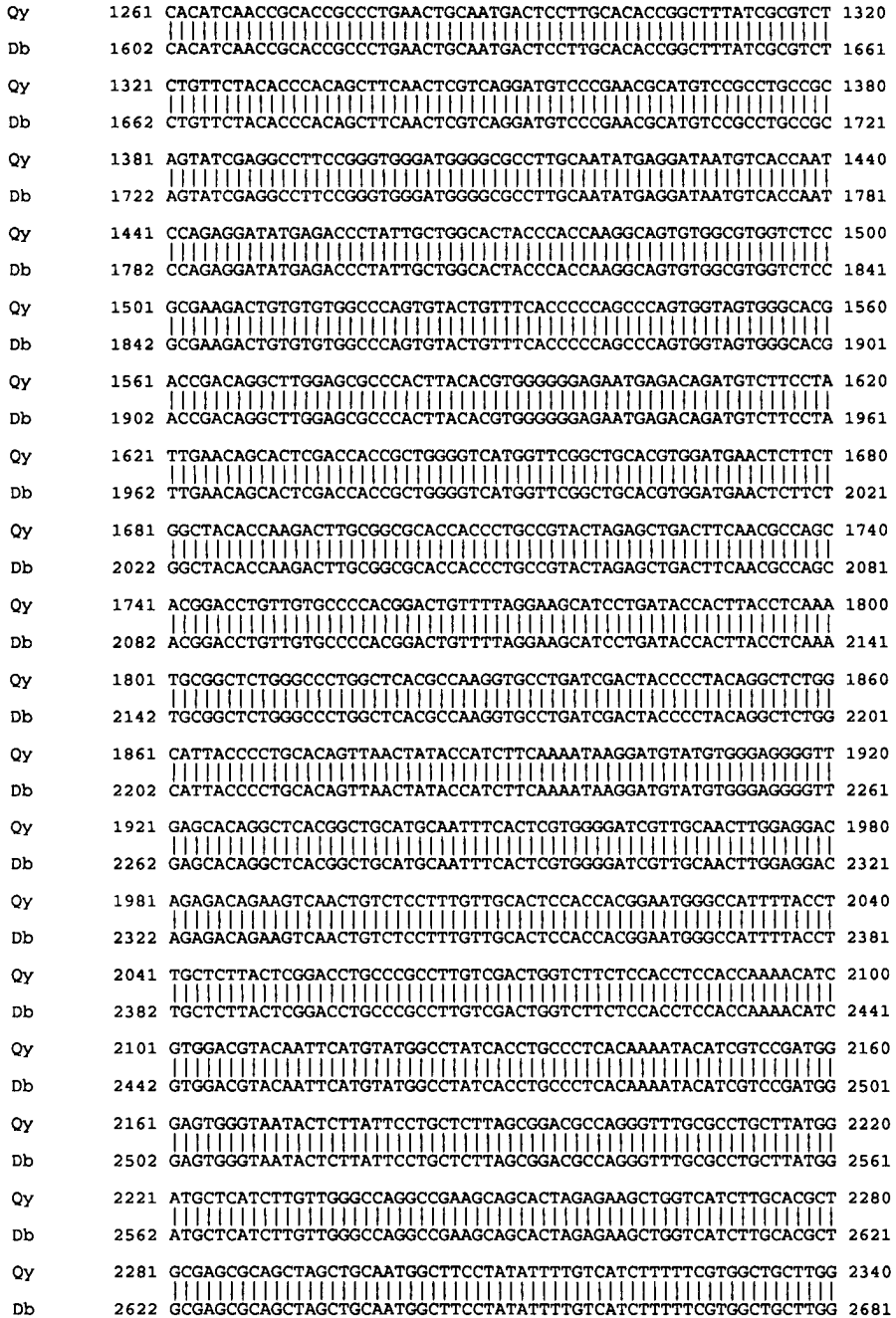
Qy        961 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1302 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1361

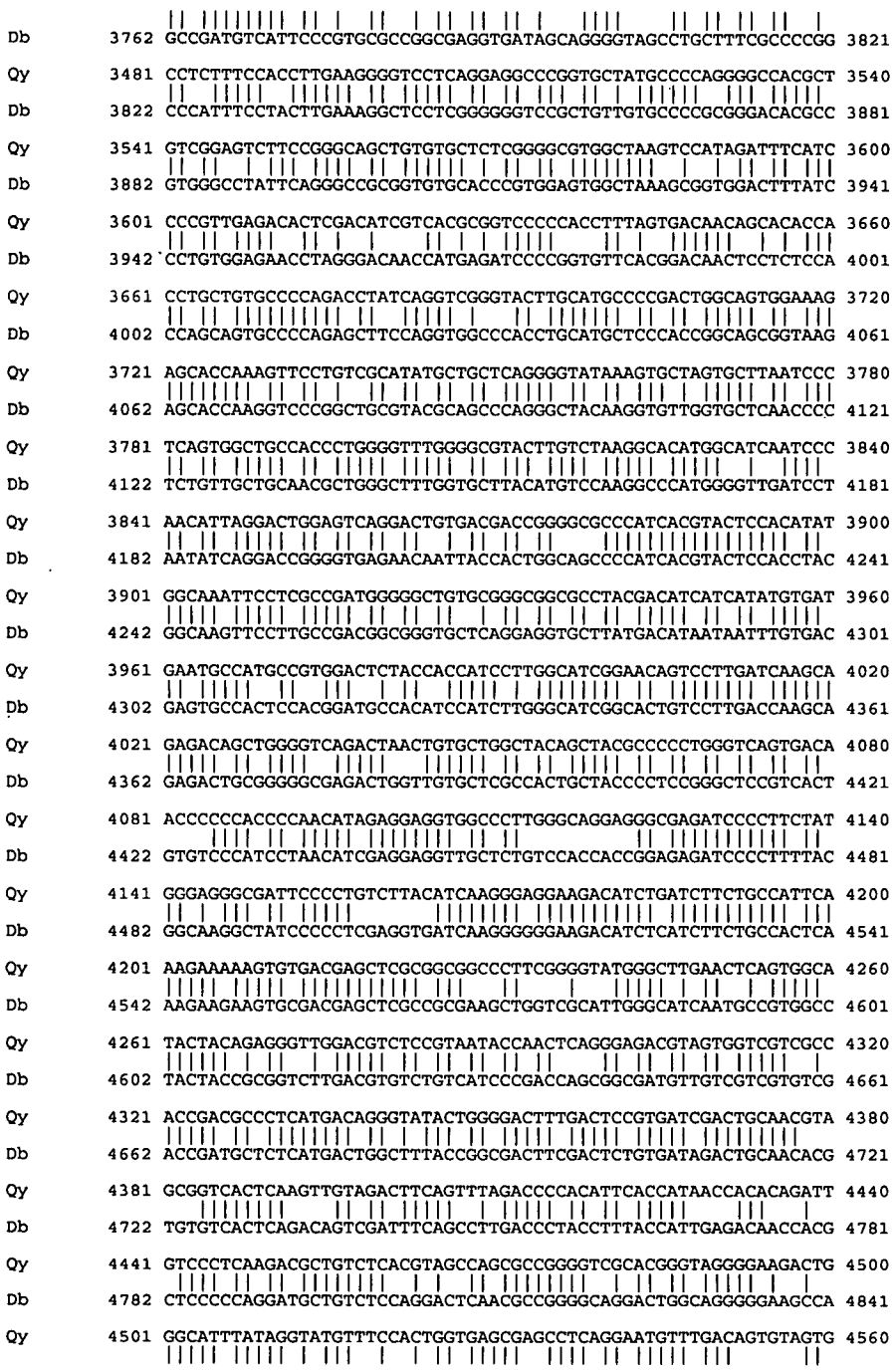
Qy       1021 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1362 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1421

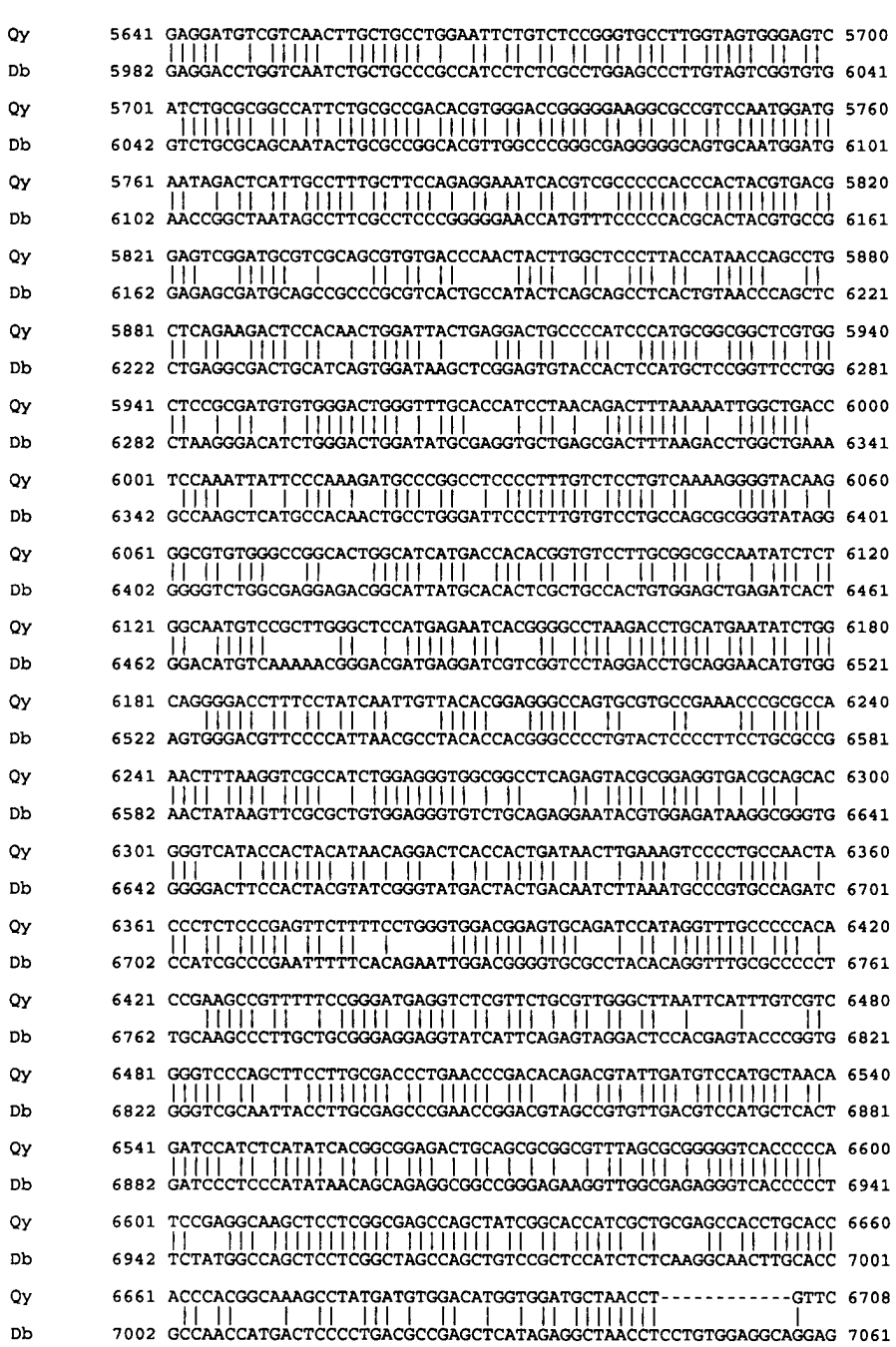
Qy       1081 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1422 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1481

Qy       1141 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1482 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1541

Qy       1201 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1542 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1601
```


Qy 1261 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1320
Db 1602 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1661
Qy 1321 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1380
Db 1662 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1721
Qy 1381 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1440
Db 1722 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1781
Qy 1441 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1500
Db 1782 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1841
Qy 1501 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1560
Db 1842 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1901
Qy 1561 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1620
Db 1902 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1961
Qy 1621 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 1680
Db 1962 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 2021
Qy 1681 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 1740
Db 2022 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 2081
Qy 1741 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 1800
Db 2082 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 2141
Qy 1801 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 1860
Db 2142 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 2201
Qy 1861 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 1920
Db 2202 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 2261
Qy 1921 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 1980
Db 2262 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 2321
Qy 1981 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2040
Db 2322 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2381
Qy 2041 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2100
Db 2382 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2441
Qy 2101 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2160
Db 2442 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2501
Qy 2161 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2220
Db 2502 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2561
Qy 2221 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2280
Db 2562 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2621
Qy 2281 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATATTTTGTCATCTTTTTCGTGGCTGCTTGG 2340
Db 2622 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATATTTTGTCATCTTTTTCGTGGCTGCTTGG 2681
Qy 2341 TACATCAAGGGTCGGGTAGTCCCCTTAGCTACCTATTCCCTCACTGGCCTGTGGTCCTTT 2400
Db 2682 TACATCAAGGGTCGGGTAGTCCCCTTAGCTACCTATTCCCTCACTGGCCTGTGGTCCTTT 2741
Qy 2401 AGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTGCATGGC 2460
Db 2742 AGCCTACTGCTCCTAGCATTGCCCCAACAGGCATATGCACTGGACACGGAGGTGGCCGCG 2801
Qy 2461 CAGATAGGAGCGGCTCTGCTGGTAATGATCACTCTCTTTACTCTCACCCCCGGGTATAAG 2520
Db 2802 TCGTGTGGCGGCGTTGTTCTTGTCGGGTTAATGGCGCTGACTCTGTCGCCATATTACAAG 2861
Qy 2521 ACCCTTCTCAGCCGGTTTTTGTGGTGGTTGTGCTATCTTCTGACCCTGGGGGAAGCTATG 2580
Db 2862 CGCTATATCAGCTGGTGCATGTGGTGGCTTCAGTATTTTCTGACCAGAGTAGAAGCGCAA 2921
Qy 2581 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGCATCATATGGGCC 2640
Db 2922 CTGCACGTGTGGGTTCCCCCCCCTCAACGTCCGGGGGGGGGCGCGATGCCGTCATCTTACTC 2981
Qy 2641 GTCGCCATATTCTACCCAGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 2700
Db 2982 ATGTGTGTAGTACACCCGACCCTGGTATTTGACATCACCAAACTACTCCTGGCCATCTTC 3041
Qy 2701 GGGCCTGCTTACCTCCTAAAAGGTGCTTTGACGCGCGTGCCGTACTTCGTCAGGGCTCAC 2760
Db 3042 GGACCCCTTTGGATTCTTCAAGCCAGTTTGCTTAAAGTCCCCTACTTCGTGCGCGTTCAA 3101
Qy 2761 GCTCTACTGAGGATGTGCACCATGGCAAGGCATCTCGCGGGGGGCAGGTACGTCCAGATG 2820
Db 3102 GGCCTTCTCCGGATCTGCGCGCTAGCGCGGAAGATAGCCGGAGGTCATTACGTGCAAATG 3161
Qy 2821 GCGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 2880
Db 3162 GCCATCATCAAGTTAGGGGCGCTTACTGGCACCTATGTGTATAACCATCTCACCCCTCTT 3221
Qy 2881 TCGGATTGGGCTGCTAGTGGCCTGCGGGACCTGGCGGTCGCCGTTGAGCCTATCATCTTC 2940
Db 3222 CGAGACTGGGCGCACAACGGCCTGCGAGATCTGGCCGTGGCTGTGGAACCAGTCGTCTTC 3281
Qy 2941 AGTCCGATGGAGAAGAAAGTCATTGTCTGGGGAGCGGAGACAGCTGCTTGTGGGGACATT 3000
Db 3282 TCCCGAATGGAGACCAAGCTCATCACGTGGGGGGCAGATACCGCCGCGTGCGGTGACATC 3341
Qy 3001 TTACACGGACTTCCCGTGTCCGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3060
Db 3342 ATCAACGGCTTGCCCGTCTCTGCCCGTAGGGGCCAGGAGATACTGCTTGGGCCAGCCGAC 3401
Qy 3061 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3120
Db 3402 GGAATGGTCTCCAAGGGGTGGAGGTTGCTGGCGCCCATCACGGCGTACGCCCAGCAGACG 3461
Qy 3121 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCT 3180
Db 3462 AGAGGCCTCCTAGGGTGTATAATCACCAGCCTGACTGGCCGGGACAAAAACCAAGTGGAG 3521
Qy 3181 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3240
Db 3522 GGTGAGGTCCAGATCGTGTCAACTGCTACCCAAACCTTCCTGGCAACGTGCATCAATGGG 3581
Qy 3241 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3300
Db 3582 GTATGCTGGACTGTCTACCACGGGGCCGGAACGAGGACCATCGCATCACCCAAGGGTCCT 3641
Qy 3301 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGG 3360
Db 3642 GTCATCCAGATGTATACCAATGTGGACCAAGACCTTGTGGGCTGGCCCGCTCCTCAAGGT 3701
Qy 3361 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3420
Db 3702 TCCCGCTCATTGACACCCTGTACCTGCGGCTCCTCGGACCTTTACCTGGTCACGAGGCAC 3761
Qy 3421 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3480


Db 3762 GCCGATGTCATTCCCGTGCGCCGGCGAGGTGATAGCAGGGGTAGCCTGCTTTCGCCCCGG 3821
Qy 3481 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3540
Db 3822 CCCATTTCCTACTTGAAAGGCTCCTCGGGGGGTCCGCTGTTGTGCCCCGCGGGACACGCC 3881
Qy 3541 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3600
Db 3882 GTGGGCCTATTCAGGGCCGCGGTGTGCACCCGTGGAGTGGCTAAAGCGGTGGACTTTATC 3941
Qy 3601 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 3660
Db 3942 CCTGTGGAGAACCTAGGGACAACCATGAGATCCCCGGTGTTCACGGACAACTCCTCTCCA 4001
Qy 3661 CCTGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 3720
Db 4002 CCAGCAGTGCCCCAGAGCTTCCAGGTGGCCCACCTGCATGCTCCCACCGGCAGCGGTAAG 4061
Qy 3721 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 3780
Db 4062 AGCACCAAGGTCCCGGCTGCGTACGCAGCCCAGGGCTACAAGGTGTTGGTGCTCAACCCC 4121
Qy 3781 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 3840
Db 4122 TCTGTTGCTGCAACGCTGGGCTTTGGTGCTTACATGTCCAAGGCCCATGGGGTTGATCCT 4181
Qy 3841 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 3900
Db 4182 AATATCAGGACCGGGGTGAGAACAATTACCACTGGCAGCCCCATCACGTACTCCACCTAC 4241
Qy 3901 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 3960
Db 4242 GGCAAGTTCCTTGCCGACGGCGGGTGCTCAGGAGGTGCTTATGACATAATAATTTGTGAC 4301
Qy 3961 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4020
Db 4302 GAGTGCCACTCCACGGATGCCACATCCATCTTGGGCATCGGCACTGTCCTTGACCAAGCA 4361
Qy 4021 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4080
Db 4362 GAGACTGCGGGGGCGAGACTGGTTGTGCTCGCCACTGCTACCCCTCCGGGCTCCGTCACT 4421
Qy 4081 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4140
Db 4422 GTGTCCCATCCTAACATCGAGGAGGTTGCTCTGTCCACCACCGGAGAGATCCCCTTTTAC 4481
Qy 4141 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4200
Db 4482 GGCAAGGCTATCCCCCTCGAGGTGATCAAGGGGGGAAGACATCTCATCTTCTGCCACTCA 4541
Qy 4201 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4260
Db 4542 AAGAAGAAGTGCGACGAGCTCGCCGCGAAGCTGGTCGCATTGGGCATCAATGCCGTGGCC 4601
Qy 4261 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4320
Db 4602 TACTACCGCGGTCTTGACGTGTCTGTCATCCCGACCAGCGGCGATGTTGTCGTCGTGTCG 4661
Qy 4321 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4380
Db 4662 ACCGATGCTCTCATGACTGGCTTTACCGGCGACTTCGACTCTGTGATAGACTGCAACACG 4721
Qy 4381 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4440
Db 4722 TGTGTCACTCAGACAGTCGATTTCAGCCTTGACCCTACCTTTACCATTGAGACAACCACG 4781
Qy 4441 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4500
Db 4782 CTCCCCCAGGATGCTGTCTCCAGGACTCAACGCCGGGGCAGGACTGGCAGGGGGAAGCCA 4841
Qy 4501 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4560
Db 4842 GGCATCTATAGATTTGTGGCACCGGGGGAGCGCCCCTCCGGCATGTTCGACTCGTCCGTC 4901
Qy 4561 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4620
Db 4902 CTCTGTGAGTGCTATGACGCGGGCTGTGCTTGGTATGAGCTCACGCCCGCCGAGACTACA 4961
Qy 4621 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 4680
Db 4962 GTTAGGCTACGAGCGTACATGAACACCCCGGGGCTTCCCGTGTGCCAGGACCATCTTGAA 5021
Qy 4681 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 4740
Db 5022 TTTTGGGAGGGCGTCTTTACGGGCCTCACTCATATAGATGCCCACTTTTTATCCCAGACA 5081
Qy 4741 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 4800
Db 5082 AAGCAGAGTGGGGAGAACTTTCCTTACCTGGTAGCGTACCAAGCCACCGTGTGCGCTAGG 5141
Qy 4801 GCCAAAGCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 4860
Db 5142 GCTCAAGCCCCTCCCCCATCGTGGGACCAGATGTGGAAGTGTTTGATCCGCCTTAAACCC 5201
Qy 4861 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 4920
Db 5202 ACCCTCCATGGGCCAACACCCCTGCTATACAGACTGGGCGCTGTTCAGAATGAAGTCACC 5261
Qy 4921 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 4980
Db 5262 CTGACGCACCCAATCACCAAATACATCATGACATGCATGTCGGCCGACCTGGAGGTCGTC 5321
Qy 4981 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5040
Db 5322 ACGAGCACCTGGGTGCTCGTTGGCGGCGTCCTGGCTGCTCTGGCCGCGTATTGCCTGTCA 5381
Qy 5041 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5100
Db 5382 ACAGGCTGCGTGGTCATAGTGGGCAGGATCGTCTTGTCCGGGAAGCCGGCAATTATACCT 5441
Qy 5101 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5160
Db 5442 GACAGGGAGGTTCTCTACCAGGAGTTCGATGAGATGGAAGAGTGCTCTCAGCACTTACCG 5501
Qy 5161 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5220
Db 5502 TACATCGAGCAAGGGATGATGCTCGCTGAGCAGTTCAAGCAGAAGGCCCTCGGCCTCCTG 5561
Qy 5221 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5280
Db 5562 CAGACCGCGTCCCGCCATGCAGAGGTTATCACCCCTGCTGTCCAGACCAACTGGCAGAAA 5621
Qy 5281 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5340
Db 5622 CTCGAGGTCTTTTGGGCGAAGCACATGTGGAATTTCATCAGTGGGATACAATACTTGGCG 5681
Qy 5341 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5400
Db 5682 GGCCTGTCAACGCTGCCTGGTAACCCCGCCATTGCTTCATTGATGGCTTTTACAGCTGCC 5741
Qy 5401 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5460
Db 5742 GTCACCAGCCCACTAACCACTGGCCAAACCCTCCTCTTCAACATATTGGGGGGGTGGGTG 5801
Qy 5461 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5520
Db 5802 GCTGCCCAGCTCGCCGCCCCCGGTGCCGCTACTGCCTTTGTGGGTGCTGGCCTAGCTGGC 5861
Qy 5521 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5580
Db 5862 GCCGCCATCGGCAGCGTTGGACTGGGGAAGGTCCTCGTGGACATTCTTGCAGGGTATGGC 5921
Qy 5581 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5640
Db 5922 GCGGGCGTGGCGGGAGCTCTTGTAGCATTCAAGATCATGAGCGGTGAGGTCCCCTCCACG 5981


Qy 5641 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 5700
Db 5982 GAGGACCTGGTCAATCTGCTGCCCGCCATCCTCTCGCCTGGAGCCCTTGTAGTCGGTGTG 6041
Qy 5701 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 5760
Db 6042 GTCTGCGCAGCAATACTGCGCCGGCACGTTGGCCCGGGCGAGGGGGCAGTGCAATGGATG 6101
Qy 5761 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 5820
Db 6102 AACCGGCTAATAGCCTTCGCCTCCCGGGGGAACCATGTTTCCCCCACGCACTACGTGCCG 6161
Qy 5821 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 5880
Db 6162 GAGAGCGATGCAGCCGCCCGCGTCACTGCCATACTCAGCAGCCTCACTGTAACCCAGCTC 6221
Qy 5881 CTCAGAAGACTCCACAACTGGATTACTGAGGACTGCCCCATCCCATGCGGCGGCTCGTGG 5940
Db 6222 CTGAGGCGACTGCATCAGTGGATAAGCTCGGAGTGTACCACTCCATGCTCCGGTTCCTGG 6281
Qy 5941 CTCCGCGATGTGTGGGACTGGGTTTGCACCATCCTAACAGACTTTAAAAATTGGCTGACC 6000
Db 6282 CTAAGGGACATCTGGGACTGGATATGCGAGGTGCTGAGCGACTTTAAGACCTGGCTGAAA 6341
Qy 6001 TCCAAATTATTCCCAAAGATGCCCGGCCTCCCCTTTGTCTCCTGTCAAAAGGGGTACAAG 6060
Db 6342 GCCAAGCTCATGCCACAACTGCCTGGGATTCCCTTTGTGTCCTGCCAGCGCGGGTATAGG 6401
Qy 6061 GGCGTGTGGGCCGGCACTGGCATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6120
Db 6402 GGGGTCTGGCGAGGAGACGGCATTATGCACACTCGCTGCCACTGTGGAGCTGAGATCACT 6461
Qy 6121 GGCAATGTCCGCTTGGGCTCCATGAGAATCACGGGGCCTAAGACCTGCATGAATATCTGG 6180
Db 6462 GGACATGTCAAAAACGGGACGATGAGGATCGTCGGTCCTAGGACCTGCAGGAACATGTGG 6521
Qy 6181 CAGGGGACCTTTCCTATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCGCCA 6240
Db 6522 AGTGGGACGTTCCCCATTAACGCCTACACCACGGGCCCCTGTACTCCCCTTCCTGCGCCG 6581
Qy 6241 AACTTTAAGGTCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6300
Db 6582 AACTATAAGTTCGCGCTGTGGAGGGTGTCTGCAGAGGAATACGTGGAGATAAGGCGGGTG 6641
Qy 6301 GGGTCATACCACTACATAACAGGACTCACCACTGATAACTTGAAAGTCCCCTGCCAACTA 6360
Db 6642 GGGGACTTCCACTACGTATCGGGTATGACTACTGACAATCTTAAATGCCCGTGCCAGATC 6701
Qy 6361 CCCTCTCCCGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCACA 6420
Db 6702 CCATCGCCCGAATTTTTCACAGAATTGGACGGGGTGCGCCTACACAGGTTTGCGCCCCCT 6761
Qy 6421 CCGAAGCCGTTTTTCCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6480
Db 6762 TGCAAGCCCTTGCTGCGGGAGGAGGTATCATTCAGAGTAGGACTCCACGAGTACCCGGTG 6821
Qy 6481 GGGTCCCAGCTTCCTTGCGACCCTGAACCCGACACAGACGTATTGATGTCCATGCTAACA 6540
Db 6822 GGGTCGCAATTACCTTGCGAGCCCGAACCGGACGTAGCCGTGTTGACGTCCATGCTCACT 6881
Qy 6541 GATCCATCTCATATCACGGCGGAGACTGCAGCGCGGCGTTTAGCGCGGGGGTCACCCCCA 6600
Db 6882 GATCCCTCCCATATAACAGCAGAGGCGGCCGGGAGAAGGTTGGCGAGAGGGTCACCCCCT 6941
Qy 6601 TCCGAGGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 6660
Db 6942 TCTATGGCCAGCTCCTCGGCTAGCCAGCTGTCCGCTCCATCTCTCAAGGCAACTTGCACC 7001
Qy 6661 ACCCACGGCAAAGCCTATGATGTGGACATGGTGGATGCTAACCT----------GTTC 6708
Db 7002 GCCAACCATGACTCCCCTGACGCCGAGCTCATAGAGGCTAACCTCCTGTGGAGGCAGGAG 7061
Qy 6709 ATGGGGGGCGATGTGACTCGGATAGAGTCTGGGTCCAAAGTGGTCGTTCTGGACTCTCTC 6768
Db 7062 ATGGGCGGCAACATCACCAGGGTTGAGTCAGAGAACAAAGTGGTGATTCTGGACTCCTTC 7121
Qy 6769 GACCCAATGGTCGAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCAGAATACATGCTC 6828
Db 7122 GATCCGCTTGTGGCAGAGGAGGATGAGCGGGAGGTCTCCGTACCTGCAGAAATTCTGCGG 7181
Qy 6829 CCCAAGAAGAGGTTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCG 6888
Db 7182 AAGTCTCGGAGATTCGCCCGGGCCCTGCCCGTCTGGGCGCGGCCGGACTACAACCCCCCG 7241
Qy 6889 CTTGTGGAATCGTGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTC 6948
Db 7242 CTAGTAGAGACGTGGAAAAAGCCTGACTACGAACCACCTGTGGTCCATGGCTGCCCGCTA 7301
Qy 6949 CCTCCTCCTAGGAAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGT 7008
Db 7302 CCACCTCCACGGTCCCCTCCTGTGCCTCCGCCTCGGAAAAAGCGTACGGTGGTCCTCACC 7361
Qy 7009 GAGGACTCCATAGGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCC 7068
Db 7362 GAATCAACCCTATCTACTGCCTTGGCCGAGCTTGCCACCAAAAGTTTTGGCAGCTCCTCA 7421
Qy 7069 CCAAGCGGCGATTCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACG 7128
Db 7422 ACTTCCGGCATTACGGGCGACAATACGACAACATCCTCTGAGCCCGCC------------ 7469
Qy 7129 CCTCCTGATGAGTTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCTCGAG 7188
Db 7470 CCTTCTGGCTGCCCCCCCGACTCCGACGTTGAGTCCTATTCTTCCATGCCCCCCCTGGAG 7529
Qy 7189 GGGGAGCTTGGAGATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCAG 7248
Db 7530 GGGGAGCCTGGGGATCCGGA-- 7549
Qy 7249 GGGGGGGTGGCAGCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGAC 7308
Db 7550 -------------TCTCAGCGACGGGTCATGGTCGACGGTCAGTAGTGGGGCCGACACG 7595
Qy 7309 GACTCCGTCGTGTGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGT 7368
Db 7596 GAAGATGTCGTGTGCTGCTCAATGTCTTATTCCTGGACAGGCGCACTCGTCACCCCGTGC 7655
Qy 7369 AGTCCCGAAGAGGAGAAGTTACCGATTAACCCCTTGAGCAACTCCCTGTTGCGATATCAC 7428
Db 7656 GCTGCGGAAGAACAAAAACTGCCCATCAACGCACTGAGCAACTCGTTGCTACGCCATCAC 7715
Qy 7429 AACAAGGTGTACTGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTT 7488
Db 7716 AATCTGGTGTATTCCACCACTTCACGCAGTGCTTGCCAAAGGCAGAAGAAAGTCACATTT 7775
Qy 7489 GATAGGATGCAAGTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCG 7548
Db 7776 GACAGACTGCAAGTTCTGGACAGCCATTACCAGGACGTGCTCAAGGAGGTCAAAGCAGCG 7835
Qy 7549 GCCTCCAAGGTCACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCC 7608
Db 7836 GCGTCAAAAGTGAAGGCTAACTTGCTATCCGTAGAGGAAGCTTGCAGCCTGACGCCCCCA 7895
Qy 7609 CATTCTGCAAGATCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGG 7668
Db 7896 CATTCAGCCAAATCCAAGTTTGGCTATGGGGCAAAAGACGTCCGTTGCCATGCCAGAAAG 7955
Qy 7669 GCCGTTAACCACATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATT 7728
Db 7956 GCCGTAGCCCACATCAACTCCGTGTGGAAAGACCTTCTGGAAGACAGTGTAACACCAATA 8015
Qy 7729 CCCACAACCATTATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAG 7788
Db 8016 GACACTACCATCATGGCCAAGAACGAGGTTTTCTGCGTTCAGCCTGAGAAGGGGGGTCGT 8075
Qy 7789 AAAGCAGCTCGCCTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCC 7848

```
Db       8076 AAGCCAGCTCGTCTCATCGTGTTCCCCGACCTGGGCGTGCGCGTGTGCGAGAAGATGGCC 8135

Qy       7849 CTTTATGACATTACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAG 7908

Db       8136 CTGTACGACGTGGTTAGCAAGCTCCCCCTGGCCGTGATGGGAAGCTCCTACGGATTCCAA 8195

Qy       7909 TATTCCCCCGCTCAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCT 7968

Db       8196 TACTCACCAGGACAGCGGGTTGAATTCCTCGTGCAAGCGTGGAAGTCCAAGAAGACCCCG 8255

Qy       7969 ATGGGTTTTTCGTATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGG 8028

Db       8256 ATGGGGTTCTCGTATGATACCCGCTGTTTTGACTCCACAGTCACTGAGAGCGACATCCGT 8315

Qy       8029 ACTGAGGAGTCCATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACAC 8088

Db       8316 ACGGAGGAGGCAATTTACCAATGTTGTGACCTGGACCCCCAAGCCCGCGTGGCCATCAAG 8375

Qy       8089 TCGCTAACTGAGAGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGC 8148

Db       8376 TCCCTCACTGAGAGGCTTTATGTTGGGGGCCCTCTTACCAATTCAAGGGGGGAAAACTGC 8435

Qy       8149 GGGTACAGGCGTTGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACA 8208

Db       8436 GGCTACCGCAGGTGCCGCGCGAGCGGCGTACTGACAACTAGCTGTGGTAACACCCTCACT 8495

Qy       8209 TGCTACGTGAAAGCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTG 8268

Db       8496 TGCTACATCAAGGCCCGGGCAGCCTGTCGAGCCGCAGGGCTCCAGGACTGCACCATGCTC 8555

Qy       8269 GTATGCGGCGATGACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGG 8328

Db       8556 GTGTGTGGCGACGACTTAGTCGTTATCTGTGAAAGTGCGGGGGTCCAGGAGGACGCGGCG 8615

Qy       8329 AACCTGAGAGCCTTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCC 8388

Db       8616 AGCCTGAGAGCCTTCACGGAGGCTATGACCAGGTACTCCGCCCCCCCCGGGGACCCCCCA 8675

Qy       8389 AGACCGGAGTATGATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTG 8448

Db       8676 CAACCAGAATACGACTTGGAGCTTATAACATCATGCTCCTCCAACGTGTCAGTCGCCCAC 8735

Qy       8449 GGCCCACAAGGCCGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGG 8508

Db       8736 GACGGCGCTGGAAAGAGGGTCTACTACCTTACCCGTGACCCTACAACCCCCCTCGCGAGA 8795

Qy       8509 GCTGCCTGGGAAACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAG 8568

Db       8796 GCCGCGTGGGAGACAGCAAGACACACTCCAGTCAATTCCTGGCTAGGCAACATAATCATG 8855

Qy       8569 TACGCCCCGACCATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATG 8628

Db       8856 TTTGCCCCCACACTGTGGGCGAGGATGATACTGATGACCCATTTCTTTAGCGTCCTCATA 8915

Qy       8629 GCTCAAGACACGCTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTG 8688

Db       8916 GCCAGGGATCAGCTTGAACAGGCTCTTAACTGTGAGATCTACGGAGCCTGCTACTCCATA 8975

Qy       8689 AGTCCCTTGGACCTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTG 8748

Db       8976 GAACCACTGGATCTACCTCCAATCATTCAAAGACTCCATGGCCTCAGCGCATTTTCACTC 9035

Qy       8749 CACACATACACTCCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCG 8808

Db       9036 CACAGTTACTCTCCAGGTGAAATCAATAGGGTGGCCGCATGCCTCAGAAAACTTGGGGTC 9095

Qy       8809 CCACCCCTCAGAGCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGT 8868

Db       9096 CCGCCCTTGCGAGCTTGGAGACACCGGGCCCGGAGCGTCCGCGCTAGGCTTCTGTCCAGA 9155

Qy       8869 GGGGGGAGAGCGGCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTC 8928

Db       9156 GGAGGCAGGGCTGCTATATGTGGCAAGTACCTCTTCAACTGGGCAGTAAGAACAAAGCTC 9215

Qy       8929 AAACTCACTCCATTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGC 8988

Db       9216 AAACTCACTCCAATAGCGGCCGCTGGCCGGCTGGACTTGTCCGGTTGGTTCACGGCTGGC 9275

Qy       8989 GCCGGCGGGGGCGACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTT 9048

Db       9276 TACAGCGGGGGAGACATTTATCACAGCGTGTCTCATGCCCGGCCCCGCTGGTTCTGGTTT 9335

Qy       9049 GGCCTACTCCTACTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCG 9098

Db       9336 TGCCTACTCCTGCTCGCTGCAGGGGTAGGCATCTACCTCCTCCCCAACCG 9385
```

RESULT 8
AAC86647
ID AAC86647 standard; DNA; 9611 BP.
XX
AC AAC86647;
XX
DT 02-APR-2001 (first entry)
XX
DE Nucleotide sequence of chimeric Hepatitis C virus clone J6S.
XX
KW HCV; HCV strain HC-J6CH; HCV genotype 2a; antiviral; vaccine; ss.
XX
OS Synthetic.
OS Hepatitis C virus.
XX
FH Key Location/Qualifiers
FT CDS 342. .9389
FT /*tag= a
XX
PN WO200075338-A2.
XX
PD 14-DEC-2000.
XX
PF 02-JUN-2000; 2000WO-US015446.
XX
PR 04-JUN-1999; 99US-0137693P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Yanagi M, Bukh J, Emerson SU, Purcell RH;
XX
DR WPI; 2001-061728/07.
DR P-PSDB; AAB30732.
XX
PT Nucleic acid molecule encoding human hepatitis C virus of genotype 2a for
PT developing vaccines, for diagnosis of hepatitis C virus and in screening
PT assays for identification of antiviral agents.
XX
PS Disclosure; Page 131-134; 167pp; English.
XX
CC AAC86645-48 represent chimeric cDNA clones of infectious Hepatitis C
CC virus (HCV). In each clone the C, E1 and E2 genes are derived from HCV
CC strain HC-J6CH genotype 2a. The p7 protein was derived from HCV strain HC
CC -J6CH or HCV strain pCV-H77C, and the NS genes were derived from HCV
CC strain pCV-H77C. Such HCV sequences are capable of expressing the virus
CC when transfected into cells. The HCV protein is useful for assaying
CC candidate antiviral agents for activity against HCV. Antibodies specific
CC for HCV polypeptide are useful in prevention and treatment of diseases
CC caused by HCV in animals, in particular humans. The HCV polypeptides
CC serve as immunogens in the development of vaccines for preventing HCV in
CC mammals or as antigens in diagnostic assays for detecting the presence of
CC HCV in biological samples. The HCV polynucleotide is also useful for
CC identifying cell lines capable of supporting the replication of HCV in
CC vitro and to produce attenuated viral strains via passage in vitro or in
CC vivo
XX
SQ Sequence 9611 BP; 1935 A; 2896 C; 2682 G; 2098 T; 0 U; 0 Other;

Query Match 59.3%; Score 5399.2; DB 5; Length 9611;
Best Local Similarity 75.0%; Pred. No. 0;
Matches 6829; Conservative 0; Mismatches 2203; Indels 78; Gaps 3;

```
Qy          1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        342 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 401

Qy         61 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        402 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 461

Qy        121 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        462 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 521

Qy        181 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        522 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 581

Qy        241 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        582 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 641

Qy        301 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        642 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 701

Qy        361 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        702 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 761

Qy        421 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        762 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 821

Qy        481 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        822 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 881

Qy        541 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        882 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 941

Qy        601 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        942 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 1001

Qy        661 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1002 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 1061

Qy        721 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1062 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 1121

Qy        781 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1122 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 1181

Qy        841 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1182 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 1241

Qy        901 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1242 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 1301

Qy        961 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1302 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1361

Qy       1021 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1362 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1421

Qy       1081 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1422 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1481

Qy       1141 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1482 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1541

Qy       1201 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1542 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1601

Qy       1261 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1602 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1661

Qy       1321 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1662 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1721

Qy       1381 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1722 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1781

Qy       1441 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1782 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1841

Qy       1501 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1842 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1901

Qy       1561 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1902 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1961

Qy       1621 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1962 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 2021

Qy       1681 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2022 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 2081

Qy       1741 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2082 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 2141

Qy       1801 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2142 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 2201

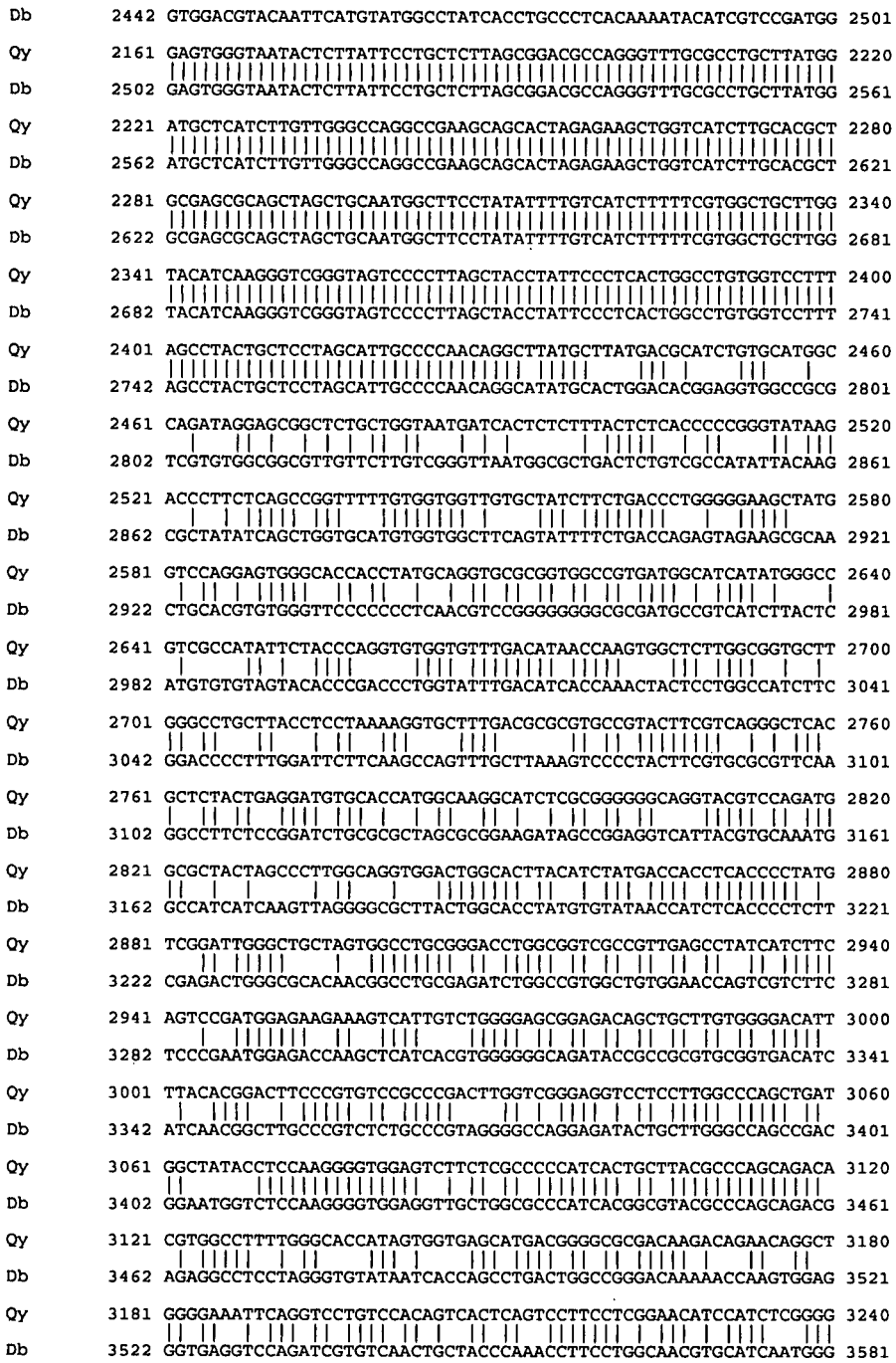
Qy       1861 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2202 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 2261

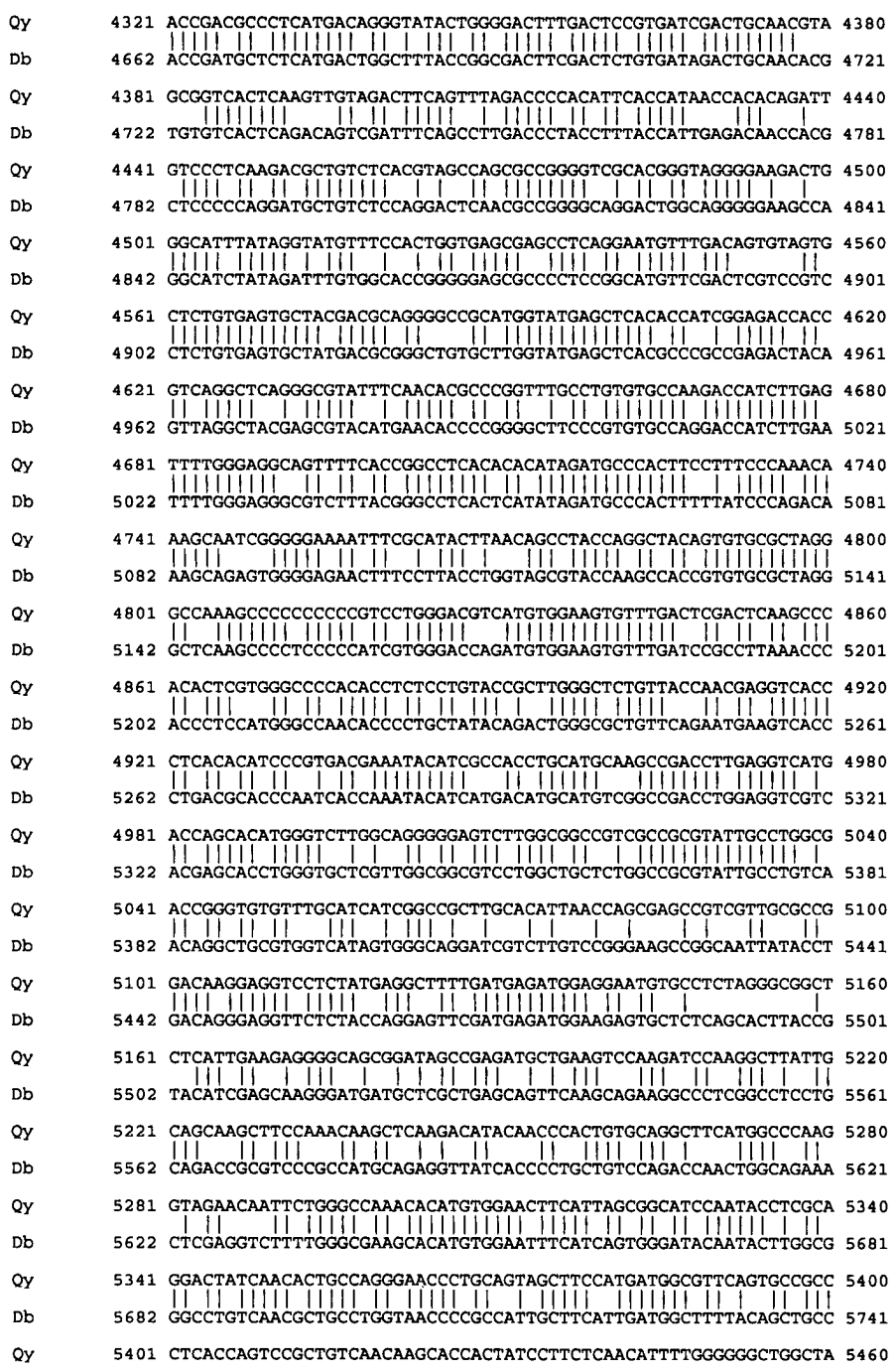
Qy       1921 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2262 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 2321

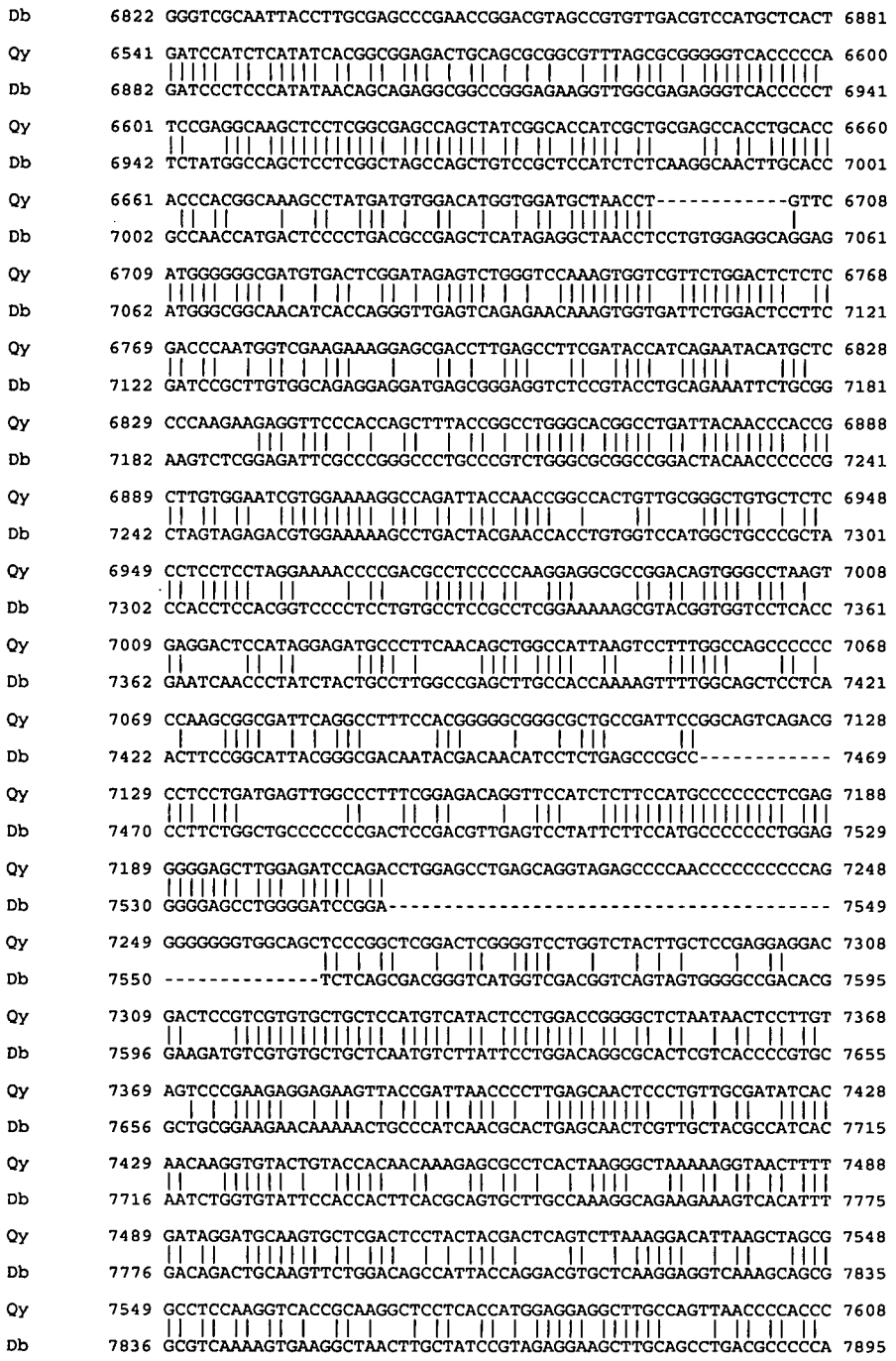
Qy       1981 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2322 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2381

Qy       2041 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2382 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2441

Qy       2101 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```


Db 2442 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2501
Qy 2161 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2220
Db 2502 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2561
Qy 2221 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2280
Db 2562 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2621
Qy 2281 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATATTTTGTCATCTTTTTCGTGGCTGCTTGG 2340
Db 2622 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATATTTTGTCATCTTTTTCGTGGCTGCTTGG 2681
Qy 2341 TACATCAAGGGTCGGGTAGTCCCCTTAGCTACCTATTCCCTCACTGGCCTGTGGTCCTTT 2400
Db 2682 TACATCAAGGGTCGGGTAGTCCCCTTAGCTACCTATTCCCTCACTGGCCTGTGGTCCTTT 2741
Qy 2401 AGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTGCATGGC 2460
Db 2742 AGCCTACTGCTCCTAGCATTGCCCCAACAGGCATATGCACTGGACACGGAGGTGGCCGCG 2801
Qy 2461 CAGATAGGAGCGGCTCTGCTGGTAATGATCACTCTCTTTACTCTCACCCCCGGGTATAAG 2520
Db 2802 TCGTGTGGCGGCGTTGTTCTTGTCGGGTTAATGGCGCTGACTCTGTCGCCATATTACAAG 2861
Qy 2521 ACCCTTCTCAGCCGGTTTTTGTGGTGGTTGTGCTATCTTCTGACCCTGGGGGAAGCTATG 2580
Db 2862 CGCTATATCAGCTGGTGCATGTGGTGGCTTCAGTATTTTCTGACCAGAGTAGAAGCGCAA 2921
Qy 2581 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGCATCATATGGGCC 2640
Db 2922 CTGCACGTGTGGGTTCCCCCCCTCAACGTCCGGGGGGGGCGCGATGCCGTCATCTTACTC 2981
Qy 2641 GTCGCCATATTCTACCCAGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 2700
Db 2982 ATGTGTGTAGTACACCCGACCCTGGTATTTGACATCACCAAACTACTCCTGGCCATCTTC 3041
Qy 2701 GGGCCTGCTTACCTCCTAAAAGGTGCTTTGACGCGCGTGCCGTACTTCGTCAGGGCTCAC 2760
Db 3042 GGACCCCTTTGGATTCTTCAAGCCAGTTTGCTTAAAGTCCCCTACTTCGTGCGCGTTCAA 3101
Qy 2761 GCTCTACTGAGGATGTGCACCATGGCAAGGCATCTCGCGGGGGGCAGGTACGTCCAGATG 2820
Db 3102 GGCCTTCTCCGGATCTGCGCGCTAGCGCGGAAGATAGCCGGAGGTCATTACGTGCAAATG 3161
Qy 2821 GCGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 2880
Db 3162 GCCATCATCAAGTTAGGGGCGCTTACTGGCACCTATGTGTATAACCATCTCACCCCTCTT 3221
Qy 2881 TCGGATTGGGCTGCTAGTGGCCTGCGGGACCTGGCGGTCGCCGTTGAGCCTATCATCTTC 2940
Db 3222 CGAGACTGGGCGCACAACGGCCTGCGAGATCTGGCCGTGGCTGTGGAACCAGTCGTCTTC 3281
Qy 2941 AGTCCGATGGAGAAGAAAGTCATTGTCTGGGGAGCGGAGACAGCTGCTTGTGGGGACATT 3000
Db 3282 TCCCGAATGGAGACCAAGCTCATCACGTGGGGGGCAGATACCGCCGCGTGCGGTGACATC 3341
Qy 3001 TTACACGGACTTCCCGTGTCCGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3060
Db 3342 ATCAACGGCTTGCCCGTCTCTGCCCGTAGGGGCCAGGAGATACTGCTTGGGCCAGCCGAC 3401
Qy 3061 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3120
Db 3402 GGAATGGTCTCCAAGGGGTGGAGGTTGCTGGCGCCCATCACGGCGTACGCCCAGCAGACG 3461
Qy 3121 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCT 3180
Db 3462 AGAGGCCTCCTAGGGTGTATAATCACCAGCCTGACTGGCCGGGACAAAAACCAAGTGGAG 3521
Qy 3181 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3240
Db 3522 GGTGAGGTCCAGATCGTGTCAACTGCTACCCAAACCTTCCTGGCAACGTGCATCAATGGG 3581
Qy 3241 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3300
Db 3582 GTATGCTGGACTGTCTACCACGGGGCCGGAACGAGGACCATCGCATCACCCAAGGGTCCT 3641
Qy 3301 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGG 3360
Db 3642 GTCATCCAGATGTATACCAATGTGGACCAAGACCTTGTGGGCTGGCCCGCTCCTCAAGGT 3701
Qy 3361 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3420
Db 3702 TCCCGCTCATTGACACCCTGTACCTGCGGCTCCTCGGACCTTTACCTGGTCACGAGGCAC 3761
Qy 3421 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3480
Db 3762 GCCGATGTCATTCCCGTGCGCCGGCGAGGTGATAGCAGGGGTAGCCTGCTTTCGCCCCGG 3821
Qy 3481 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3540
Db 3822 CCCATTTCCTACTTGAAAGGCTCCTCGGGGGGTCCGCTGTTGTGCCCCGCGGGACACGCC 3881
Qy 3541 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3600
Db 3882 GTGGGCCTATTCAGGGCCGCGGTGTGCACCCGTGGAGTGGCTAAAGCGGTGGACTTTATC 3941
Qy 3601 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 3660
Db 3942 CCTGTGGAGAACCTAGGGACAACCATGAGATCCCCGGTGTTCACGGACAACTCCTCTCCA 4001
Qy 3661 CCTGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 3720
Db 4002 CCAGCAGTGCCCCAGAGCTTCCAGGTGGCCCACCTGCATGCTCCCACCGGCAGCGGTAAG 4061
Qy 3721 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 3780
Db 4062 AGCACCAAGGTCCCGGCTGCGTACGCAGCCCAGGGCTACAAGGTGTTGGTGCTCAACCCC 4121
Qy 3781 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 3840
Db 4122 TCTGTTGCTGCAACGCTGGGCTTTGGTGCTTACATGTCCAAGGCCCATGGGGTTGATCCT 4181
Qy 3841 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 3900
Db 4182 AATATCAGGACCGGGGGTGAGAACAATTACCACTGGCAGCCCCATCACGTACTCCACCTAC 4241
Qy 3901 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 3960
Db 4242 GGCAAGTTCCTTGCCGACGGCGGGTGCTCAGGAGGTGCTTATGACATAATAATTTGTGAC 4301
Qy 3961 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4020
Db 4302 GAGTGCCACTCCACGGATGCCACATCCATCTTGGGCATCGGCACTGTCCTTGACCAAGCA 4361
Qy 4021 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4080
Db 4362 GAGACTGCGGGGGCGAGACTGGTTGTGCTCGCCACTGCTACCCCTCCGGGCTCCGTCACT 4421
Qy 4081 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4140
Db 4422 GTGTCCCATCCTAACATCGAGGAGGTTGCTCTGTCCACCACCGGAGAGATCCCCTTTTAC 4481
Qy 4141 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4200
Db 4482 GGCAAGGCTATCCCCCTCGAGGTGATCAAGGGGGGAAGACATCTCATCTTCTGCCACTCA 4541
Qy 4201 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4260
Db 4542 AAGAAGAAGTGCGACGAGCTCGCCGCGAAGCTGGTCGCATTGGGCATCAATGCCGTGGCC 4601
Qy 4261 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4320
Db 4602 TACTACCGCGGTCTTGACGTGTCTGTCATCCCGACCAGCGGCGATGTTGTCGTCGTGTCG 4661


Qy 4321 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4380
Db 4662 ACCGATGCTCTCATGACTGGCTTTACCGGCGACTTCGACTCTGTGATAGACTGCAACACG 4721
Qy 4381 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4440
Db 4722 TGTGTCACTCAGACAGTCGATTTCAGCCTTGACCCTACCTTTACCATTGAGACAACCACG 4781
Qy 4441 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4500
Db 4782 CTCCCCCAGGATGCTGTCTCCAGGACTCAACGCCGGGGCAGGACTGGCAGGGGGAAGCCA 4841
Qy 4501 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4560
Db 4842 GGCATCTATAGATTTGTGGCACCGGGGGAGCGCCCCTCCGGCATGTTCGACTCGTCCGTC 4901
Qy 4561 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4620
Db 4902 CTCTGTGAGTGCTATGACGCGGGCTGTGCTTGGTATGAGCTCACGCCCGCCGAGACTACA 4961
Qy 4621 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 4680
Db 4962 GTTAGGCTACGAGCGTACATGAACACCCCGGGGCTTCCCGTGTGCCAGGACCATCTTGAA 5021
Qy 4681 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 4740
Db 5022 TTTTGGGAGGGCGTCTTTACGGGCCTCACTCATATAGATGCCCACTTTTTATCCCAGACA 5081
Qy 4741 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 4800
Db 5082 AAGCAGAGTGGGGAGAACTTTCCTTACCTGGTAGCGTACCAAGCCACCGTGTGCGCTAGG 5141
Qy 4801 GCCAAAGCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 4860
Db 5142 GCTCAAGCCCCTCCCCCATCGTGGGACCAGATGTGGAAGTGTTTGATCCGCCTTAAACCC 5201
Qy 4861 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 4920
Db 5202 ACCCTCCATGGGCCAACACCCCTGCTATACAGACTGGGCGCTGTTCAGAATGAAGTCACC 5261
Qy 4921 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 4980
Db 5262 CTGACGCACCCAATCACCAAATACATCATGACATGCATGTCGGCCGACCTGGAGGTCGTC 5321
Qy 4981 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5040
Db 5322 ACGAGCACCTGGGTGCTCGTTGGCGGCGTCCTGGCTGCTCTGGCCGCGTATTGCCTGTCA 5381
Qy 5041 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5100
Db 5382 ACAGGCTGCGTGGTCATAGTGGGCAGGATCGTCTTGTCCGGGAAGCCGGCAATTATACCT 5441
Qy 5101 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5160
Db 5442 GACAGGGAGGTTCTCTACCAGGAGTTCGATGAGATGGAAGAGTGCTCTCAGCACTTACCG 5501
Qy 5161 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5220
Db 5502 TACATCGAGCAAGGGATGATGCTCGCTGAGCAGTTCAAGCAGAAGGCCCTCGGCCTCCTG 5561
Qy 5221 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5280
Db 5562 CAGACCGCGTCCCGCCATGCAGAGGTTATCACCCCTGCTGTCCAGACCAACTGGCAGAAA 5621
Qy 5281 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5340
Db 5622 CTCGAGGTCTTTTGGGCGAAGCACATGTGGAATTTCATCAGTGGGATACAATACTTGGCG 5681
Qy 5341 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5400
Db 5682 GGCCTGTCAACGCTGCCTGGTAACCCCGCCATTGCTTCATTGATGGCTTTTACAGCTGCC 5741
Qy 5401 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5460
Db 5742 GTCACCAGCCCACTAACCACTGGCCAAACCCTCCTCTTCAACATATTGGGGGGGTGGGTG 5801
Qy 5461 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5520
Db 5802 GCTGCCCAGCTCGCCGCCCCCGGTGCCGCTACTGCCTTTGTGGGTGCTGGCCTAGCTGGC 5861
Qy 5521 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5580
Db 5862 GCCGCCATCGGCAGCGTTGGACTGGGGAAGGTCCTCGTGGACATTCTTGCAGGGTATGGC 5921
Qy 5581 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5640
Db 5922 GCGGGCGTGGCGGGAGCTCTTGTAGCATTCAAGATCATGAGCGGTGAGGTCCCCTCCACG 5981
Qy 5641 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 5700
Db 5982 GAGGACCTGGTCAATCTGCTGCCCGCCATCCTCTCGCCTGGAGCCCTTGTAGTCGGTGTG 6041
Qy 5701 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 5760
Db 6042 GTCTGCGCAGCAATACTGCGCCGGCACGTTGGCCCGGGCGAGGGGGCAGTGCAATGGATG 6101
Qy 5761 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 5820
Db 6102 AACCGGCTAATAGCCTTCGCCTCCCGGGGGAACCATGTTTCCCCCACGCACTACGTGCCG 6161
Qy 5821 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 5880
Db 6162 GAGAGCGATGCAGCCGCCCGCGTCACTGCCATACTCAGCAGCCTCACTGTAACCCAGCTC 6221
Qy 5881 CTCAGAAGACTCCACAACTGGATTACTGAGGACTGCCCCATCCCATGCGGCGGCTCGTGG 5940
Db 6222 CTGAGGCGACTGCATCAGTGGATAAGCTCGGAGTGTACCACTCCATGCTCCGGTTCCTGG 6281
Qy 5941 CTCCGCGATGTGTGGGACTGGGTTTGCACCATCCTAACAGACTTTAAAAATTGGCTGACC 6000
Db 6282 CTAAGGGACATCTGGGACTGGATATGCGAGGTGCTGAGCGACTTTAAGACCTGGCTGAAA 6341
Qy 6001 TCCAAATTATTCCCAAAGATGCCCGGCCTCCCCTTTGTCTCCTGTCAAAAGGGGTACAAG 6060
Db 6342 GCCAAGCTCATGCCACAACTGCCTGGGATTCCCTTTGTGTCCTGCCAGCGCGGGTATAGG 6401
Qy 6061 GGCGTGTGGGCCGGCACTGGCATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6120
Db 6402 GGGGTCTGGCGAGGAGACGGCATTATGCACACTCGCTGCCACTGTGGAGCTGAGATCACT 6461
Qy 6121 GGCAATGTCCGCTTGGGCTCCATGAGAATCACGGGGCCTAAGACCTGCATGAATATCTGG 6180
Db 6462 GGACATGTCAAAAACGGGACGATGAGGATCGTCGGTCCTAGGACCTGCAGGAACATGTGG 6521
Qy 6181 CAGGGGACCTTTCCTATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCGCCA 6240
Db 6522 AGTGGGACGTTCCCCATTAACGCCTACACCACGGGCCCCTGTACTCCCCTTCCTGCGCCG 6581
Qy 6241 AACTTTAAGGTCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6300
Db 6582 AACTATAAGTTCGCGCTGTGGAGGGTGTCTGCAGAGGAATACGTGGAGATAAGGCGGGTG 6641
Qy 6301 GGGTCATACCACTACATAACAGGACTCACCACTGATAACTTGAAAGTCCCCTGCCAACTA 6360
Db 6642 GGGGACTTCCACTACGTATCGGGTATGACTACTGACAATCTTAAATGCCCGTGCCAGATC 6701
Qy 6361 CCCTCTCCCGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCACA 6420
Db 6702 CCATCGCCCGAATTTTTCACAGAATTGGACGGGGTGCGCCTACACAGGTTTGCGCCCCCT 6761
Qy 6421 CCGAAGCCGTTTTTCCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6480
Db 6762 TGCAAGCCCTTGCTGCGGGAGGAGGTATCATTCAGAGTAGGACTCCACGAGTACCCGGTG 6821
Qy 6481 GGGTCCCAGCTTCCTTGCGACCCTGAACCCGACACAGACGTATTGATGTCCATGCTAACA 6540


Db 6822 GGGTCGCAATTACCTTGCGAGCCCGAACCGGACGTAGCCGTGTTGACGTCCATGCTCACT 6881
Qy 6541 GATCCATCTCATATCACGGCGGAGACTGCAGCGCGGCGTTTAGCGCGGGGGTCACCCCCA 6600
Db 6882 GATCCCTCCCATATAACAGCAGAGGCGGCCGGGAGAAGGTTGGCGAGAGGGTCACCCCCT 6941
Qy 6601 TCCGAGGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 6660
Db 6942 TCTATGGCCAGCTCCTCGGCTAGCCAGCTGTCCGCTCCATCTCTCAAGGCAACTTGCACC 7001
Qy 6661 ACCCACGGCAAAGCCTATGATGTGGACATGGTGGATGCTAACCT----------GTTC 6708
Db 7002 GCCAACCATGACTCCCCTGACGCCGAGCTCATAGAGGCTAACCTCCTGTGGAGGCAGGAG 7061
Qy 6709 ATGGGGGGCGATGTGACTCGGATAGAGTCTGGGTCCAAAGTGGTCGTTCTGGACTCTCTC 6768
Db 7062 ATGGGCGGCAACATCACCAGGGTTGAGTCAGAGAACAAAGTGGTGATTCTGGACTCCTTC 7121
Qy 6769 GACCCAATGGTCGAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCAGAATACATGCTC 6828
Db 7122 GATCCGCTTGTGGCAGAGGAGGATGAGCGGGAGGTCTCCGTACCTGCAGAAATTCTGCGG 7181
Qy 6829 CCCAAGAAGAGGTTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCG 6888
Db 7182 AAGTCTCGGAGATTCGCCCGGGCCCTGCCCGTCTGGGCGCGGCCGGACTACAACCCCCCG 7241
Qy 6889 CTTGTGGAATCGTGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTC 6948
Db 7242 CTAGTAGAGACGTGGAAAAAGCCTGACTACGAACCACCTGTGGTCCATGGCTGCCCGCTA 7301
Qy 6949 CCTCCTCCTAGGAAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGT 7008
Db 7302 CCACCTCCACGGTCCCCTCCTGTGCCTCCGCCTCGGAAAAAGCGTACGGTGGTCCTCACC 7361
Qy 7009 GAGGACTCCATAGGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCC 7068
Db 7362 GAATCAACCCTATCTACTGCCTTGGCCGAGCTTGCCACCAAAAGTTTTGGCAGCTCCTCA 7421
Qy 7069 CCAAGCGGCGATTCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACG 7128
Db 7422 ACTTCCGGCATTACGGGCGACAATACGACAACATCCTCTGAGCCCGCC---------- 7469
Qy 7129 CCTCCTGATGAGTTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCTCGAG 7188
Db 7470 CCTTCTGGCTGCCCCCCCGACTCCGACGTTGAGTCCTATTCTTCCATGCCCCCCCTGGAG 7529
Qy 7189 GGGGAGCTTGGAGATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCAG 7248
Db 7530 GGGGAGCCTGGGGATCCGGA---------- 7549
Qy 7249 GGGGGGGTGGCAGCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGAC 7308
Db 7550 ----------TCTCAGCGACGGGTCATGGTCGACGGTCAGTAGTGGGGCCGACACG 7595
Qy 7309 GACTCCGTCGTGTGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGT 7368
Db 7596 GAAGATGTCGTGTGCTGCTCAATGTCTTATTCCTGGACAGGCGCACTCGTCACCCCGTGC 7655
Qy 7369 AGTCCCGAAGAGGAGAAGTTACCGATTAACCCCTTGAGCAACTCCCTGTTGCGATATCAC 7428
Db 7656 GCTGCGGAAGAACAAAAACTGCCCATCAACGCACTGAGCAACTCGTTGCTACGCCATCAC 7715
Qy 7429 AACAAGGTGTACTGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTT 7488
Db 7716 AATCTGGTGTATTCCACCACTTCACGCAGTGCTTGCCAAAGGCAGAAGAAAGTCACATTT 7775
Qy 7489 GATAGGATGCAAGTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCG 7548
Db 7776 GACAGACTGCAAGTTCTGGACAGCCATTACCAGGACGTGCTCAAGGAGGTCAAAGCAGCG 7835
Qy 7549 GCCTCCAAGGTCACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCC 7608
Db 7836 GCGTCAAAAGTGAAGGCTAACTTGCTATCCGTAGAGGAAGCTTGCAGCCTGACGCCCCCA 7895
Qy 7609 CATTCTGCAAGATCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGG 7668
Db 7896 CATTCAGCCAAATCCAAGTTTGGCTATGGGGCAAAAGACGTCCGTTGCCATGCCAGAAAG 7955
Qy 7669 GCCGTTAACCACATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATT 7728
Db 7956 GCCGTAGCCCACATCAACTCCGTGTGGAAAGACCTTCTGGAAGACAGTGTAACACCAATA 8015
Qy 7729 CCCACAACCATTATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAG 7788
Db 8016 GACACTACCATCATGGCCAAGAACGAGGTTTTCTGCGTTCAGCCTGAGAAGGGGGGTCGT 8075
Qy 7789 AAAGCAGCTCGCCTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCC 7848
Db 8076 AAGCCAGCTCGTCTCATCGTGTTCCCCGACCTGGGCGTGCGCGTGTGCGAGAAGATGGCC 8135
Qy 7849 CTTTATGACATTACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAG 7908
Db 8136 CTGTACGACGTGGTTAGCAAGCTCCCCCTGGCCGTGATGGGAAGCTCCTACGGATTCCAA 8195
Qy 7909 TATTCCCCCGCTCAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCT 7968
Db 8196 TACTCACCAGGACAGCGGGTTGAATTCCTCGTGCAAGCGTGGAAGTCCAAGAAGACCCCG 8255
Qy 7969 ATGGGTTTTTCGTATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGG 8028
Db 8256 ATGGGGTTCTCGTATGATACCCGCTGTTTTGACTCCACAGTCACTGAGAGCGACATCCGT 8315
Qy 8029 ACTGAGGAGTCCATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACAC 8088
Db 8316 ACGGAGGAGGCAATTTACCAATGTTGTGACCTGGACCCCCAAGCCCGCGTGGCCATCAAG 8375
Qy 8089 TCGCTAACTGAGAGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGC 8148
Db 8376 TCCCTCACTGAGAGGCTTTATGTTGGGGGCCCTCTTACCAATTCAAGGGGGGAAAACTGC 8435
Qy 8149 GGGTACAGGCGTTGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACA 8208
Db 8436 GGCTACCGCAGGTGCCGCGCGAGCGGCGTACTGACAACTAGCTGTGGTAACACCCTCACT 8495
Qy 8209 TGCTACGTGAAAGCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTG 8268
Db 8496 TGCTACATCAAGGCCCGGGCAGCCTGTCGAGCCGCAGGGCTCCAGGACTGCACCATGCTC 8555
Qy 8269 GTATGCGGCGATGACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGG 8328
Db 8556 GTGTGTGGCGACGACTTAGTCGTTATCTGTGAAAGTGCGGGGGTCCAGGAGGACGCGGCG 8615
Qy 8329 AACCTGAGAGCCTTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCC 8388
Db 8616 AGCCTGAGAGCCTTCACGGAGGCTATGACCAGGTACTCCGCCCCCCCCGGGGACCCCCCA 8675
Qy 8389 AGACCGGAGTATGATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTG 8448
Db 8676 CAACCAGAATACGACTTGGAGCTTATAACATCATGCTCCTCCAACGTGTCAGTCGCCCAC 8735
Qy 8449 GGCCCACAAGGCCGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGG 8508
Db 8736 GACGGCGCTGGAAAGAGGGTCTACTACCTTACCCGTGACCCTACAACCCCCCTCGCGAGA 8795
Qy 8509 GCTGCCTGGGAAACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAG 8568
Db 8796 GCCGCGTGGGAGACAGCAAGACACACTCCAGTCAATTCCTGGCTAGGCAACATAATCATG 8855
Qy 8569 TACGCCCCGACCATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATG 8628
Db 8856 TTTGCCCCCACACTGTGGGCGAGGATGATACTGATGACCCATTTCTTTAGCGTCCTCATA 8915
Qy 8629 GCTCAAGACACGCTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTG 8688
Db 8916 GCCAGGGATCAGCTTGAACAGGCTCTTAACTGTGAGATCTACGGAGCCTGCTACTCCATA 8975

```
Qy       8689 AGTCCCTTGGACCTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTG 8748
              ||  |||| || ||  | || ||| ||||  | ||||| ||   ||| ||||| ||
Db       8976 GAACCACTGGATCTACCTCCAATCATTCAAAGACTCCATGGCCTCAGCGCATTTTCACTC 9035

Qy       8749 CACACATACACTCCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCG 8808
              ||||  || ||||    ||| | |   |||||||  ||  ||||||||||||||||
Db       9036 CACAGTTACTCTCCAGGTGAAATCAATAGGGTGGCCGCATGCCTCAGAAAACTTGGGGTC 9095

Qy       8809 CCACCCCTCAGAGCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGT 8868
              || ||| |  |||| ||||    |||||| ||    ||| | ||    ||  | ||| |
Db       9096 CCGCCCTTGCGAGCTTGGAGACACCGGGCCCGGAGCGTCCGCGCTAGGCTTCTGTCCAGA 9155

Qy       8869 GGGGGGAGAGCGGCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTC 8928
              || || || || ||  | || ||    || |||||||| ||||| || |  || ||||||
Db       9156 GGAGGCAGGGCTGCTATATGTGGCAAGTACCTCTTCAACTGGGCAGTAAGAACAAAGCTC 9215

Qy       8929 AAACTCACTCCATTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGC 8988
              |||||||||||| |  |||  ||  |||  ||||| || ||| | |||||||| |  |||
Db       9216 AAACTCACTCCAATAGCGGCCGCTGGCCGGCTGGACTTGTCCGGTTGGTTCACGGCTGGC 9275

Qy       8989 GCCGGCGGGGGCGACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTT 9048
                | |||||||  |||||||||||||||||||| | ||||| |||||||  ||    |||
Db       9276 TACAGCGGGGGAGACATTTATCACAGCGTGTCTCATGCCCGGCCCCGCTGGTTCTGGTTT 9335

Qy       9049 GGCCTACTCCTACTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCG 9098
              ||||||||||  ||   || |||||||||| | | ||| ||||||   ||
Db       9336 TGCCTACTCCTGCTCGCTGCAGGGGTAGGCATCTACCTCCTCCCCAACCG 9385


RESULT 9
AAC86646
ID   AAC86646 standard; DNA; 9611 BP.
XX
AC   AAC86646;
XX
DT   02-APR-2001  (first entry)
XX
DE   Nucleotide sequence of chimeric Hepatitis C virus clone pH77(p7)-CV-J6S.
XX
KW   HCV; HCV strain HC-J6CH; HCV genotype 2a; antiviral; vaccine; ss.
XX
OS   Synthetic.
OS   Hepatitis C virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             342. .9389
FT                   /*tag=  a
XX
PN   WO200075338-A2.
XX
PD   14-DEC-2000.
XX
PF   02-JUN-2000; 2000WO-US015446.
XX
PR   04-JUN-1999;   99US-0137693P.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Yanagi M,  Bukh J,  Emerson SU,  Purcell RH;
XX
DR   WPI; 2001-061728/07.
DR   P-PSDB; AAB30731.
XX
PT   Nucleic acid molecule encoding human hepatitis C virus of genotype 2a for
PT   developing vaccines, for diagnosis of hepatitis C virus and in screening
PT   assays for identification of antiviral agents.
XX
PS   Disclosure; Page 115-118; 167pp; English.
XX
CC   AAC86645-48 represent chimeric cDNA clones of infectious Hepatitis C
CC   virus (HCV). In each clone the C, E1 and E2 genes are derived from HCV
CC   strain HC-J6CH genotype 2a. The p7 protein was derived from HCV strain HC
CC   -J6CH or HCV strain pCV-H77C, and the NS genes were derived from HCV
CC   strain pCV-H77C. Such HCV sequences are capable of expressing the virus
CC   when transfected into cells. The HCV protein is useful for assaying
CC   candidate antiviral agents for activity against HCV. Antibodies specific
CC   for HCV polypeptide are useful in prevention and treatment of diseases
CC   caused by HCV in animals, in particular humans. The HCV polypeptides
CC   serve as immunogens in the development of vaccines for preventing HCV in
CC   mammals or as antigens in diagnostic assays for detecting the presence of
CC   HCV in biological samples. The HCV polynucleotide is also useful for
CC   identifying cell lines capable of supporting the replication of HCV in
CC   vitro and to produce attenuated viral strains via passage in vitro or in
CC   vivo
XX
SQ   Sequence 9611 BP; 1924 A; 2898 C; 2694 G; 2095 T; 0 U; 0 Other;

  Query Match           58.0%;  Score 5274.4;  DB 5;  Length 9611;
  Best Local Similarity  74.1%;  Pred. No. 0;
  Matches 6751;  Conservative    0;  Mismatches 2281;  Indels   78;  Gaps    3;

Qy          1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        342 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 401

Qy         61 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        402 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 461

Qy        121 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        462 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 521

Qy        181 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        522 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 581

Qy        241 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        582 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 641

Qy        301 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        642 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 701

Qy        361 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        702 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 761

Qy        421 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        762 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 821

Qy        481 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        822 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 881

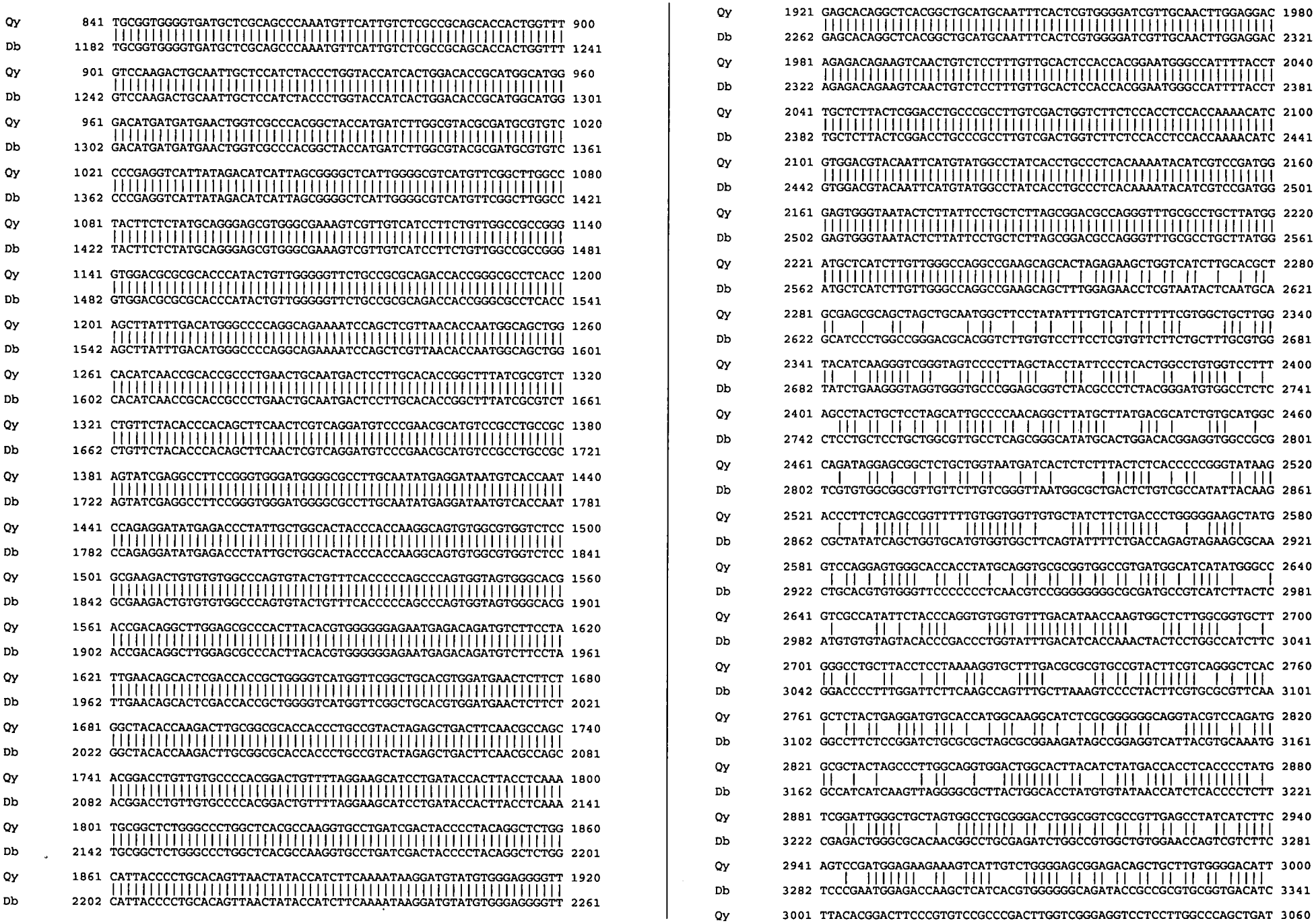
Qy        541 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        882 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 941

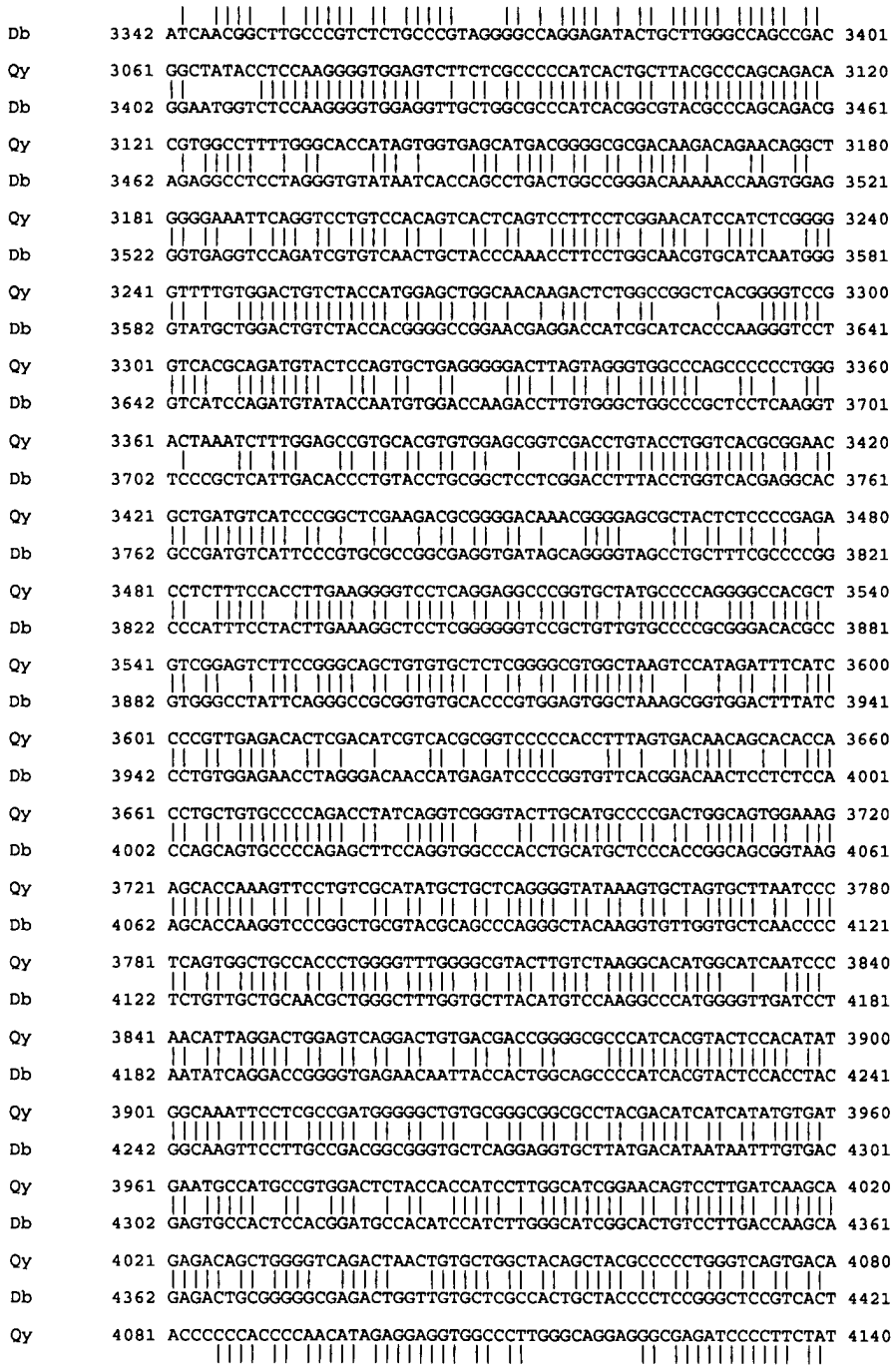
Qy        601 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        942 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 1001

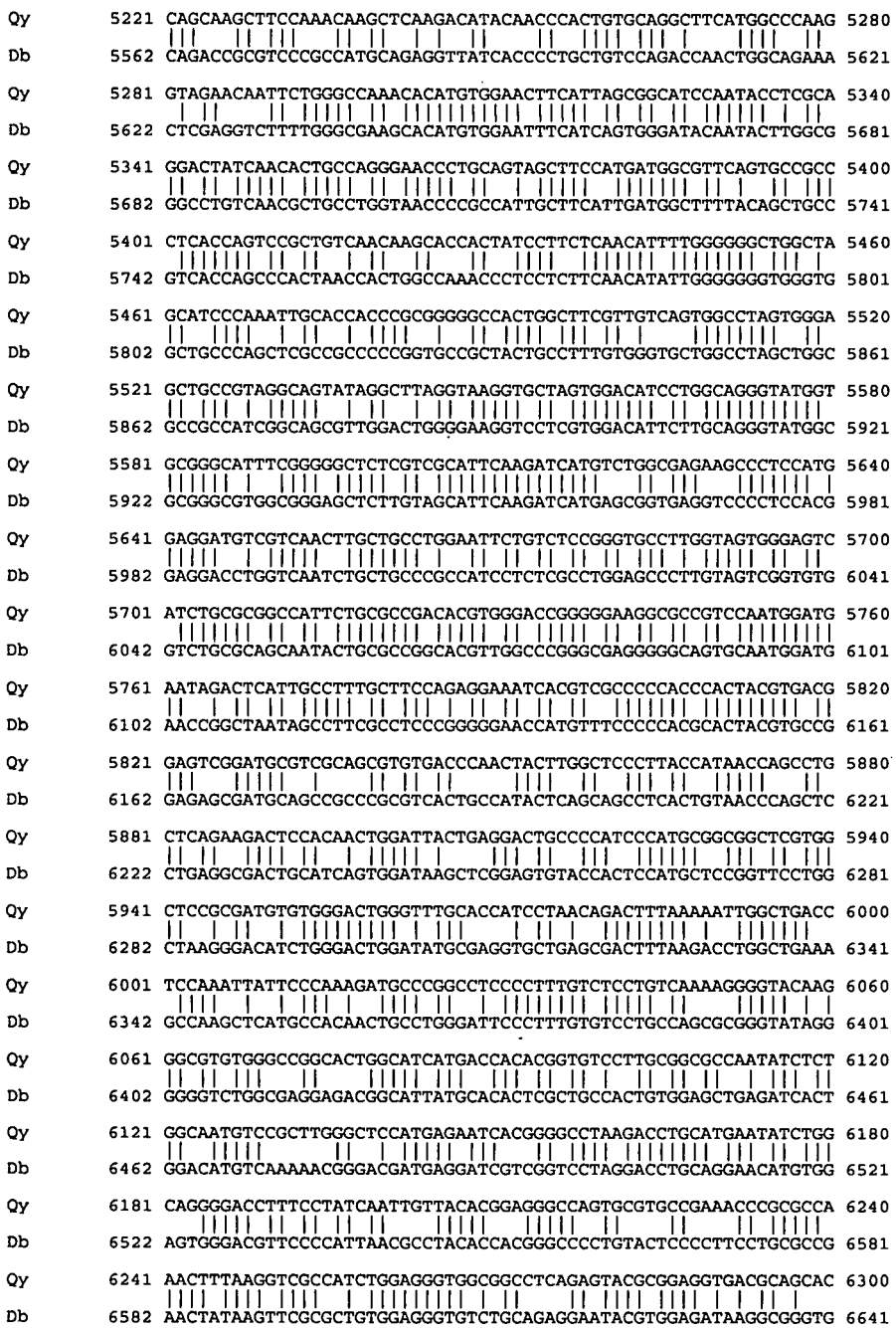
Qy        661 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1002 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 1061

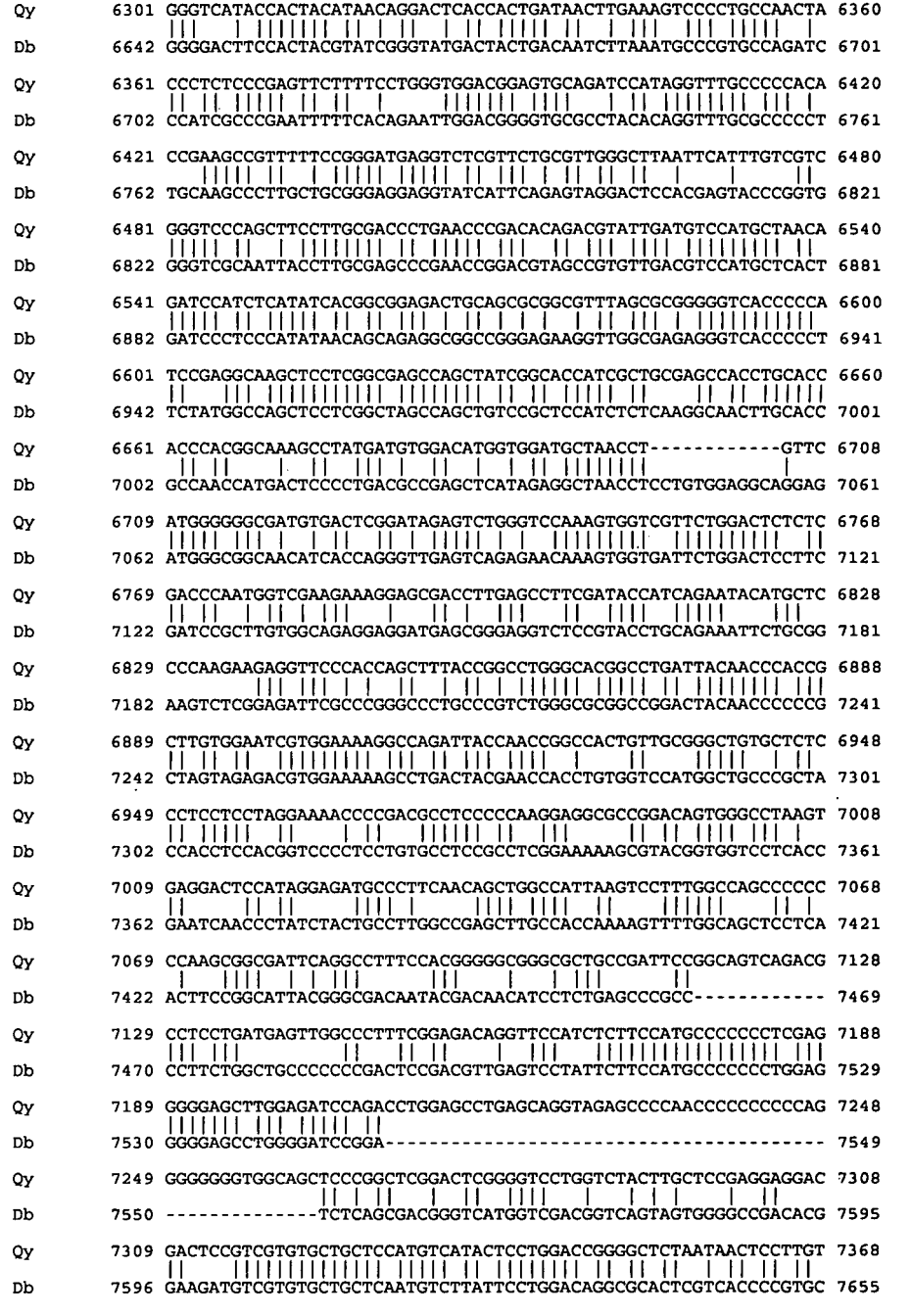
Qy        721 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1062 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 1121

Qy        781 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1122 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 1181
```


Qy 841 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 900
Db 1182 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 1241
Qy 901 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 960
Db 1242 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 1301
Qy 961 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1020
Db 1302 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1361
Qy 1021 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1080
Db 1362 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1421
Qy 1081 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1140
Db 1422 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1481
Qy 1141 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1200
Db 1482 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1541
Qy 1201 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1260
Db 1542 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1601
Qy 1261 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1320
Db 1602 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1661
Qy 1321 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1380
Db 1662 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1721
Qy 1381 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1440
Db 1722 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1781
Qy 1441 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1500
Db 1782 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1841
Qy 1501 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1560
Db 1842 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1901
Qy 1561 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1620
Db 1902 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1961
Qy 1621 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 1680
Db 1962 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 2021
Qy 1681 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 1740
Db 2022 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 2081
Qy 1741 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 1800
Db 2082 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 2141
Qy 1801 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 1860
Db 2142 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 2201
Qy 1861 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 1920
Db 2202 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 2261
Qy 1921 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 1980
Db 2262 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 2321
Qy 1981 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2040
Db 2322 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2381
Qy 2041 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2100
Db 2382 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2441
Qy 2101 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2160
Db 2442 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2501
Qy 2161 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2220
Db 2502 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2561
Qy 2221 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2280
Db 2562 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCTTTGGAGAACCTCGTAATACTCAATGCA 2621
Qy 2281 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATATTTTGTCATCTTTTTCGTGGCTGCTTGG 2340
Db 2622 GCATCCCTGGCCGGGACGCACGGTCTTGTGTCCTTCCTCGTGTTCTTCTGCTTTGCGTGG 2681
Qy 2341 TACATCAAGGGTCGGGTAGTCCCCTTAGCTACCTATTCCCTCACTGGCCTGTGGTCCTTT 2400
Db 2682 TATCTGAAGGGTAGGTGGGTGCCCGGAGCGGTCTACGCCCTCTACGGGATGTGGCCTCTC 2741
Qy 2401 AGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTGCATGGC 2460
Db 2742 CTCCTGCTCCTGCTGGCGTTGCCTCAGCGGGCATATGCACTGGACACGGAGGTGGCCGCG 2801
Qy 2461 CAGATAGGAGCGGCTCTGCTGGTAATGATCACTCTCTTTACTCTCACCCCCGGGTATAAG 2520
Db 2802 TCGTGTGGCGGCGTTGTTCTTGTCGGGTTAATGGCGCTGACTCTGTCGCCATATTACAAG 2861
Qy 2521 ACCCTTCTCAGCCGGTTTTTGTGGTGGTTGTGCTATCTTCTGACCCTGGGGGAAGCTATG 2580
Db 2862 CGCTATATCAGCTGGTGCATGTGGTGGCTTCAGTATTTTCTGACCAGAGTAGAAGCGCAA 2921
Qy 2581 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGCATCATATGGGCC 2640
Db 2922 CTGCACGTGTGGGTTCCCCCCCTCAACGTCCGGGGGGGGCGCGATGCCGTCATCTTACTC 2981
Qy 2641 GTCGCCATATTCTACCCAGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 2700
Db 2982 ATGTGTGTAGTACACCCGACCCTGGTATTTGACATCACCAAACTACTCCTGGCCATCTTC 3041
Qy 2701 GGGCCTGCTTACCTCCTAAAAGGTGCTTTGACGCGCGTGCCGTACTTCGTCAGGGCTCAC 2760
Db 3042 GGACCCCTTTGGATTCTTCAAGCCAGTTTGCTTAAAGTCCCCTACTTCGTGCGCGTTCAA 3101
Qy 2761 GCTCTACTGAGGATGTGCACCATGGCAAGGCATCTCGCGGGGGGCAGGTACGTCCAGATG 2820
Db 3102 GGCCTTCTCCGGATCTGCGCGCTAGCGCGGAAGATAGCCGGAGGTCATTACGTGCAAATG 3161
Qy 2821 GCGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 2880
Db 3162 GCCATCATCAAGTTAGGGGCGCTTACTGGCACCTATGTGTATAACCATCTCACCCCTCTT 3221
Qy 2881 TCGGATTGGGCTGCTAGTGGCCTGCGGGACCTGGCGGTCGCCGTTGAGCCTATCATCTTC 2940
Db 3222 CGAGACTGGGCGCACAACGGCCTGCGAGATCTGGCCGTGGCTGTGGAACCAGTCGTCTTC 3281
Qy 2941 AGTCCGATGGAGAAGAAAGTCATTGTCTGGGGAGCGGAGACAGCTGCTTGTGGGGACATT 3000
Db 3282 TCCCGAATGGAGACCAAGCTCATCACGTGGGGGGCAGATACCGCCGCGTGCGGTGACATC 3341
Qy 3001 TTACACGGACTTCCCGTGTCCGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3060


Db 3342 ATCAACGGCTTGCCCGTCTCTGCCCGTAGGGGCCAGGAGATACTGCTTGGGCCAGCCGAC 3401
Qy 3061 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3120
Db 3402 GGAATGGTCTCCAAGGGGTGGAGGTTGCTGGCGCCCATCACGGCGTACGCCCAGCAGACG 3461
Qy 3121 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCT 3180
Db 3462 AGAGGCCTCCTAGGGTGTATAATCACCAGCCTGACTGGCCGGGACAAAAACCAAGTGGAG 3521
Qy 3181 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3240
Db 3522 GGTGAGGTCCAGATCGTGTCAACTGCTACCCAAACCTTCCTGGCAACGTGCATCAATGGG 3581
Qy 3241 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3300
Db 3582 GTATGCTGGACTGTCTACCACGGGGCCGGAACGAGGACCATCGCATCACCCAAGGGTCCT 3641
Qy 3301 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGG 3360
Db 3642 GTCATCCAGATGTATACCAATGTGGACCAAGACCTTGTGGGCTGGCCCGCTCCTCAAGGT 3701
Qy 3361 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3420
Db 3702 TCCCGCTCATTGACACCCTGTACCTGCGGCTCCTCGGACCTTTACCTGGTCACGAGGCAC 3761
Qy 3421 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3480
Db 3762 GCCGATGTCATTCCCGTGCGCCGGCGAGGTGATAGCAGGGGTAGCCTGCTTTCGCCCCGG 3821
Qy 3481 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3540
Db 3822 CCCATTTCCTACTTGAAAGGCTCCTCGGGGGGTCCGCTGTTGTGCCCCGCGGGACACGCC 3881
Qy 3541 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3600
Db 3882 GTGGGCCTATTCAGGGCCGCGGTGTGCACCCGTGGAGTGGCTAAAGCGGTGGACTTTATC 3941
Qy 3601 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 3660
Db 3942 CCTGTGGAGAACCTAGGGACAACCATGAGATCCCCGGTGTTCACGGACAACTCCTCTCCA 4001
Qy 3661 CCTGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 3720
Db 4002 CCAGCAGTGCCCCAGAGCTTCCAGGTGGCCCACCTGCATGCTCCCACCGGCAGCGGTAAG 4061
Qy 3721 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 3780
Db 4062 AGCACCAAGGTCCCGGCTGCGTACGCAGCCCAGGGCTACAAGGTGTTGGTGCTCAACCCC 4121
Qy 3781 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 3840
Db 4122 TCTGTTGCTGCAACGCTGGGCTTTGGTGCTTACATGTCCAAGGCCCATGGGGTTGATCCT 4181
Qy 3841 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 3900
Db 4182 AATATCAGGACCGGGGTGAGAACAATTACCACTGGCAGCCCCATCACGTACTCCACCTAC 4241
Qy 3901 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 3960
Db 4242 GGCAAGTTCCTTGCCGACGGCGGGTGCTCAGGAGGTGCTTATGACATAATAATTTGTGAC 4301
Qy 3961 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4020
Db 4302 GAGTGCCACTCCACGGATGCCACATCCATCTTGGGCATCGGCACTGTCCTTGACCAAGCA 4361
Qy 4021 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4080
Db 4362 GAGACTGCGGGGGCGAGACTGGTTGTGCTCGCCACTGCTACCCCTCCGGGCTCCGTCACT 4421
Qy 4081 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4140
Db 4422 GTGTCCCATCCTAACATCGAGGAGGTTGCTCTGTCCACCACCGGAGAGATCCCCTTTTAC 4481
Qy 4141 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4200
Db 4482 GGCAAGGCTATCCCCCTCGAGGTGATCAAGGGGGGAAGACATCTCATCTTCTGCCACTCA 4541
Qy 4201 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4260
Db 4542 AAGAAGAAGTGCGACGAGCTCGCCGCGAAGCTGGTCGCATTGGGCATCAATGCCGTGGCC 4601
Qy 4261 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4320
Db 4602 TACTACCGCGGTCTTGACGTGTCTGTCATCCCGACCAGCGGCGATGTTGTCGTCGTGTCG 4661
Qy 4321 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4380
Db 4662 ACCGATGCTCTCATGACTGGCTTTACCGGCGACTTCGACTCTGTGATAGACTGCAACACG 4721
Qy 4381 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4440
Db 4722 TGTGTCACTCAGACAGTCGATTTCAGCCTTGACCCTACCTTTACCATTGAGACAACCACG 4781
Qy 4441 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4500
Db 4782 CTCCCCCAGGATGCTGTCTCCAGGACTCAACGCCGGGGCAGGACTGGCAGGGGGAAGCCA 4841
Qy 4501 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4560
Db 4842 GGCATCTATAGATTTGTGGCACCGGGGGAGCGCCCCTCCGGCATGTTCGACTCGTCCGTC 4901
Qy 4561 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4620
Db 4902 CTCTGTGAGTGCTATGACGCGGGCTGTGCTTGGTATGAGCTCACGCCCGCCGAGACTACA 4961
Qy 4621 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 4680
Db 4962 GTTAGGCTACGAGCGTACATGAACACCCCGGGGCTTCCCGTGTGCCAGGACCATCTTGAA 5021
Qy 4681 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 4740
Db 5022 TTTTGGGAGGGCGTCTTTACGGGCCTCACTCATATAGATGCCCACTTTTTATCCCAGACA 5081
Qy 4741 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 4800
Db 5082 AAGCAGAGTGGGGAGAACTTTCCTTACCTGGTAGCGTACCAAGCCACCGTGTGCGCTAGG 5141
Qy 4801 GCCAAAGCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 4860
Db 5142 GCTCAAGCCCCTCCCCCATCGTGGGACCAGATGTGGAAGTGTTTGATCCGCCTTAAACCC 5201
Qy 4861 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 4920
Db 5202 ACCCTCCATGGGCCAACACCCCTGCTATACAGACTGGGCGCTGTTCAGAATGAAGTCACC 5261
Qy 4921 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 4980
Db 5262 CTGACGCACCCAATCACCAAATACATCATGACATGCATGTCGGCCGACCTGGAGGTCGTC 5321
Qy 4981 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5040
Db 5322 ACGAGCACCTGGGTGCTCGTTGGCGGCGTCCTGGCTGCTCTGGCCGCGTATTGCCTGTCA 5381
Qy 5041 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5100
Db 5382 ACAGGCTGCGTGGTCATAGTGGGCAGGATCGTCTTGTCCGGGAAGCCGGCAATTATACCT 5441
Qy 5101 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5160
Db 5442 GACAGGGAGGTTCTCTACCAGGAGTTCGATGAGATGGAAGAGTGCTCTCAGCACTTACCG 5501
Qy 5161 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5220
Db 5502 TACATCGAGCAAGGGATGATGCTCGCTGAGCAGTTCAAGCAGAAGGCCCTCGGCCTCCTG 5561





```
Db       7656 GCTGCGGAAGAACAAAAACTGCCCATCAACGCACTGAGCAACTCGTTGCTACGCCATCAC 7715

Qy       7429 AACAAGGTGTACTGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTT 7488

Db       7716 AATCTGGTGTATTCCACCACTTCACGCAGTGCTTGCCAAAGGCAGAAGAAAGTCACATTT 7775

Qy       7489 GATAGGATGCAAGTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCG 7548

Db       7776 GACAGACTGCAAGTTCTGGACAGCCATTACCAGGACGTGCTCAAGGAGGTCAAAGCAGCG 7835

Qy       7549 GCCTCCAAGGTCACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCC 7608

Db       7836 GCGTCAAAAGTGAAGGCTAACTTGCTATCCGTAGAGGAAGCTTGCAGCCTGACGCCCCCA 7895

Qy       7609 CATTCTGCAAGATCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGG 7668

Db       7896 CATTCAGCCAAATCCAAGTTTGGCTATGGGGCAAAAGACGTCCGTTGCCATGCCAGAAAG 7955

Qy       7669 GCCGTTAACCACATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATT 7728

Db       7956 GCCGTAGCCCACATCAACTCCGTGTGGAAAGACCTTCTGGAAGACAGTGTAACACCAATA 8015

Qy       7729 CCCACAACCATTATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAG 7788

Db       8016 GACACTACCATCATGGCCAAGAACGAGGTTTTCTGCGTTCAGCCTGAGAAGGGGGGTCGT 8075

Qy       7789 AAAGCAGCTCGCCTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCC 7848

Db       8076 AAGCCAGCTCGTCTCATCGTGTTCCCCGACCTGGGCGTGCGCGTGTGCGAGAAGATGGCC 8135

Qy       7849 CTTTATGACATTACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAG 7908

Db       8136 CTGTACGACGTGGTTAGCAAGCTCCCCCTGGCCGTGATGGGAAGCTCCTACGGATTCCAA 8195

Qy       7909 TATTCCCCCGCTCAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCT 7968

Db       8196 TACTCACCAGGACAGCGGGTTGAATTCCTCGTGCAAGCGTGGAAGTCCAAGAAGACCCCG 8255

Qy       7969 ATGGGTTTTTCGTATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGG 8028

Db       8256 ATGGGGTTCTCGTATGATACCCGCTGTTTTGACTCCACAGTCACTGAGAGCGACATCCGT 8315

Qy       8029 ACTGAGGAGTCCATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACAC 8088

Db       8316 ACGGAGGAGGCAATTTACCAATGTTGTGACCTGGACCCCCAAGCCCGCGTGGCCATCAAG 8375

Qy       8089 TCGCTAACTGAGAGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGC 8148

Db       8376 TCCCTCACTGAGAGGCTTTATGTTGGGGGCCCTCTTACCAATTCAAGGGGGGAAAACTGC 8435

Qy       8149 GGGTACAGGCGTTGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACA 8208

Db       8436 GGCTACCGCAGGTGCCGCGCGAGCGGCGTACTGACAACTAGCTGTGGTAACACCCTCACT 8495

Qy       8209 TGCTACGTGAAAGCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTG 8268

Db       8496 TGCTACATCAAGGCCCGGGCAGCCTGTCGAGCCGCAGGGCTCCAGGACTGCACCATGCTC 8555

Qy       8269 GTATGCGGCGATGACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGG 8328

Db       8556 GTGTGTGGCGACGACTTAGTCGTTATCTGTGAAAGTGCGGGGGTCCAGGAGGACGCGGCG 8615

Qy       8329 AACCTGAGAGCCTTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCC 8388

Db       8616 AGCCTGAGAGCCTTCACGGAGGCTATGACCAGGTACTCCGCCCCCCCCGGGGACCCCCCA 8675

Qy       8389 AGACCGGAGTATGATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTG 8448

Db       8676 CAACCAGAATACGACTTGGAGCTTATAACATCATGCTCCTCCAACGTGTCAGTCGCCCAC 8735

Qy       8449 GGCCCACAAGGCCGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGG 8508

Db       8736 GACGGCGCTGGAAAGAGGGTCTACTACCTTACCCGTGACCCTACAACCCCCCTCGCGAGA 8795

Qy       8509 GCTGCCTGGGAAACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAG 8568

Db       8796 GCCGCGTGGGAGACAGCAAGACACACTCCAGTCAATTCCTGGCTAGGCAACATAATCATG 8855

Qy       8569 TACGCCCCGACCATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATG 8628

Db       8856 TTTGCCCCCACACTGTGGGCGAGGATGATACTGATGACCCATTTCTTTAGCGTCCTCATA 8915

Qy       8629 GCTCAAGACACGCTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTG 8688

Db       8916 GCCAGGGATCAGCTTGAACAGGCTCTTAACTGTGAGATCTACGGAGCCTGCTACTCCATA 8975

Qy       8689 AGTCCCTTGGACCTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTG 8748

Db       8976 GAACCACTGGATCTACCTCCAATCATTCAAAGACTCCATGGCCTCAGCGCATTTTCACTC 9035

Qy       8749 CACACATACACTCCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCG 8808

Db       9036 CACAGTTACTCTCCAGGTGAAATCAATAGGGTGGCCGCATGCCTCAGAAAACTTGGGGTC 9095

Qy       8809 CCACCCCTCAGAGCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGT 8868

Db       9096 CCGCCCTTGCGAGCTTGGAGACACCGGGCCCGGAGCGTCCGCGCTAGGCTTCTGTCCAGA 9155

Qy       8869 GGGGGGAGAGCGGCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTC 8928

Db       9156 GGAGGCAGGGCTGCTATATGTGGCAAGTACCTCTTCAACTGGGCAGTAAGAACAAAGCTC 9215

Qy       8929 AAACTCACTCCATTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGC 8988

Db       9216 AAACTCACTCCAATAGCGGCCGCTGGCCGGCTGGACTTGTCCGGTTGGTTCACGGCTGGC 9275

Qy       8989 GCCGGCGGGGGCGACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTT 9048

Db       9276 TACAGCGGGGGAGACATTTATCACAGCGTGTCTCATGCCCGGCCCCGCTGGTTCTGGTTT 9335

Qy       9049 GGCCTACTCCTACTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCG 9098

Db       9336 TGCCTACTCCTGCTCGCTGCAGGGGTAGGCATCTACCTCCTCCCCAACCG 9385
```

```
RESULT 10
AAC86648
ID   AAC86648 standard; DNA; 9611 BP.
XX
AC   AAC86648;
XX
DT   02-APR-2001  (first entry)
XX
DE   Nucleotide sequence of chimeric Hepatitis C virus clone pH77(p7)-J6S.
XX
KW   HCV; HCV strain HC-J6CH; HCV genotype 2a; antiviral; vaccine; ss.
XX
OS   Synthetic.
OS   Hepatitis C virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             342. .9389
FT                   /*tag=  a
XX
PN   WO200075338-A2.
XX
PD   14-DEC-2000.
XX
PF   02-JUN-2000; 2000WO-US015446.
XX
PR   04-JUN-1999;   99US-0137693P.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Yanagi M,  Bukh J,  Emerson SU,  Purcell RH;
```

XX
DR WPI; 2001-061728/07.
DR P-PSDB; AAB30733.
XX
PT Nucleic acid molecule encoding human hepatitis C virus of genotype 2a for
PT developing vaccines, for diagnosis of hepatitis C virus and in screening
PT assays for identification of antiviral agents.
XX
PS Disclosure; Page 146-149; 167pp; English.
XX
CC AAC86645-48 represent chimeric cDNA clones of infectious Hepatitis C
CC virus (HCV). In each clone the C, E1 and E2 genes are derived from HCV
CC strain HC-J6CH genotype 2a. The p7 protein was derived from HCV strain HC
CC -J6CH or HCV strain pCV-H77C, and the NS genes were derived from HCV
CC strain pCV-H77C. Such HCV sequences are capable of expressing the virus
CC when transfected into cells. The HCV protein is useful for assaying
CC candidate antiviral agents for activity against HCV. Antibodies specific
CC for HCV polypeptide are useful in prevention and treatment of diseases
CC caused by HCV in animals, in particular humans. The HCV polypeptides
CC serve as immunogens in the development of vaccines for preventing HCV in
CC mammals or as antigens in diagnostic assays for detecting the presence of
CC HCV in biological samples. The HCV polynucleotide is also useful for
CC identifying cell lines capable of supporting the replication of HCV in
CC vitro and to produce attenuated viral strains via passage in vitro or in
CC vivo
XX
SQ Sequence 9611 BP; 1925 A; 2897 C; 2695 G; 2094 T; 0 U; 0 Other;

Query Match 58.0%; Score 5274.4; DB 5; Length 9611;
Best Local Similarity 74.1%; Pred. No. 0;
Matches 6751; Conservative 0; Mismatches 2281; Indels 78; Gaps 3;

```
Qy          1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        342 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 401

Qy         61 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        402 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 461

Qy        121 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        462 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 521

Qy        181 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        522 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 581

Qy        241 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        582 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 641

Qy        301 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        642 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 701

Qy        361 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        702 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 761

Qy        421 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        762 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 821

Qy        481 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        822 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 881

Qy        541 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        882 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 941

Qy        601 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        942 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 1001

Qy        661 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1002 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 1061

Qy        721 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1062 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 1121

Qy        781 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1122 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 1181

Qy        841 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1182 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 1241

Qy        901 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1242 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 1301

Qy        961 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1302 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1361

Qy       1021 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1362 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1421

Qy       1081 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1422 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1481

Qy       1141 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1482 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1541

Qy       1201 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1542 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1601

Qy       1261 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1602 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1661

Qy       1321 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1662 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1721

Qy       1381 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1722 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1781

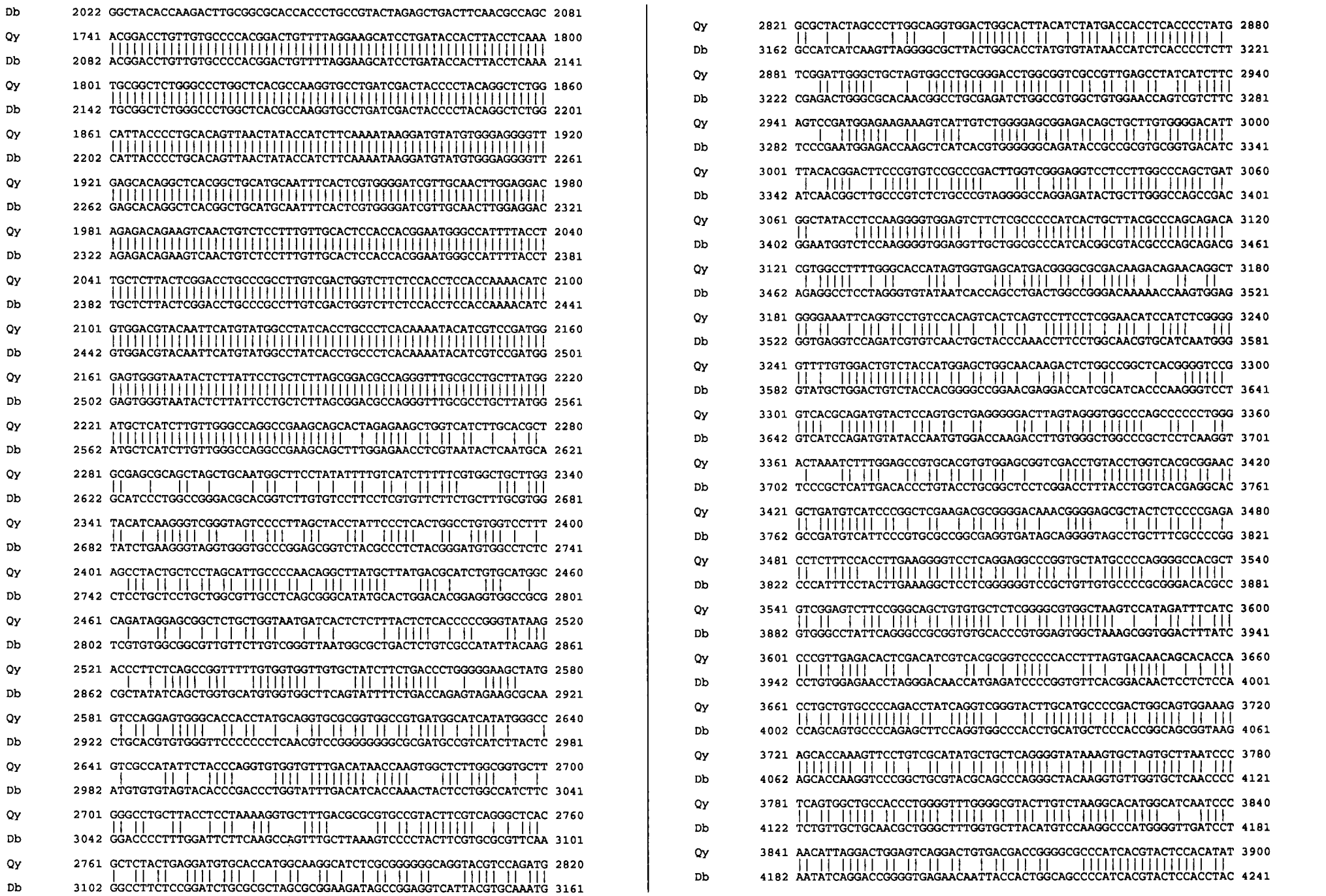
Qy       1441 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1782 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1841

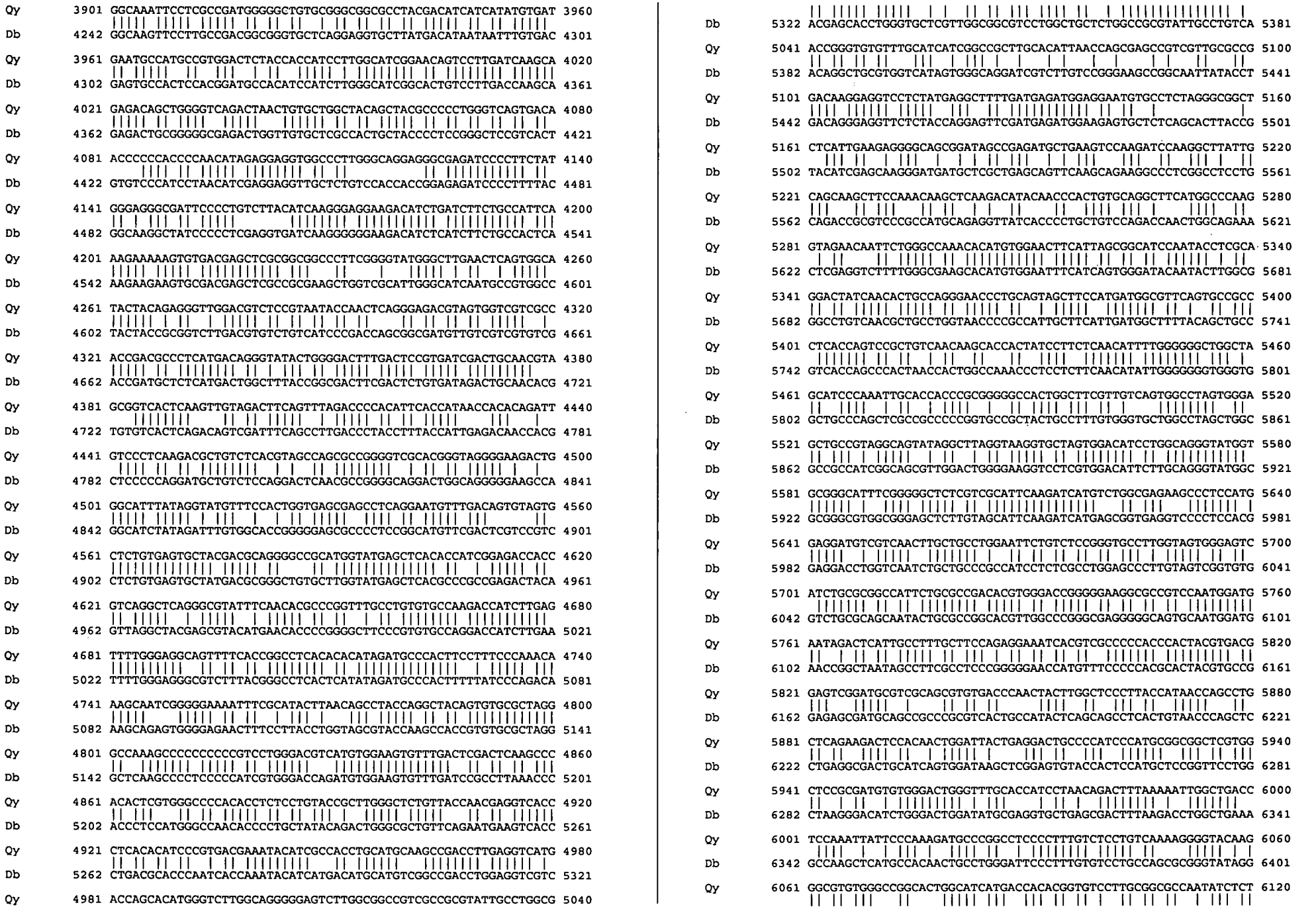
Qy       1501 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1842 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1901

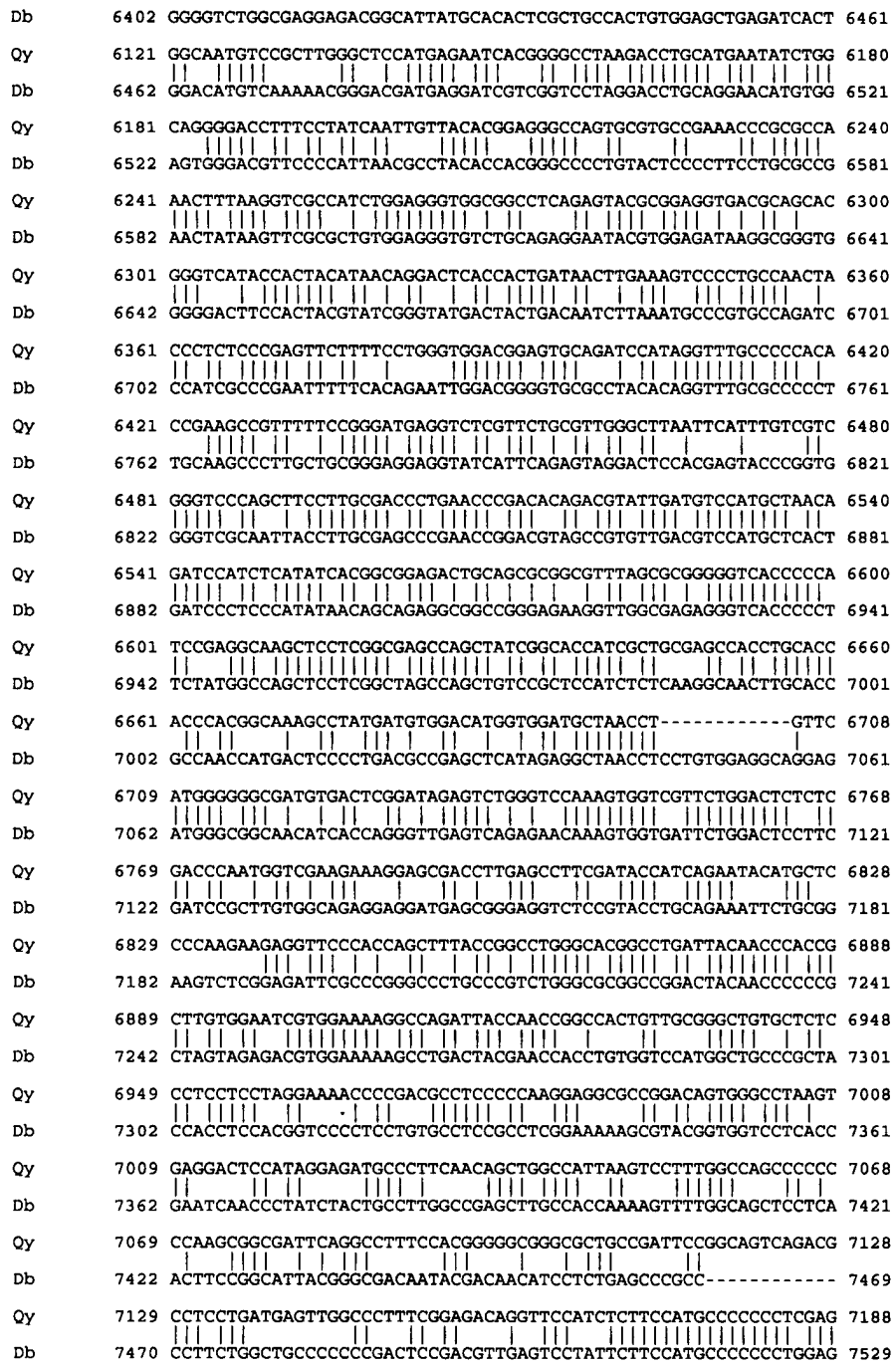
Qy       1561 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1902 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1961

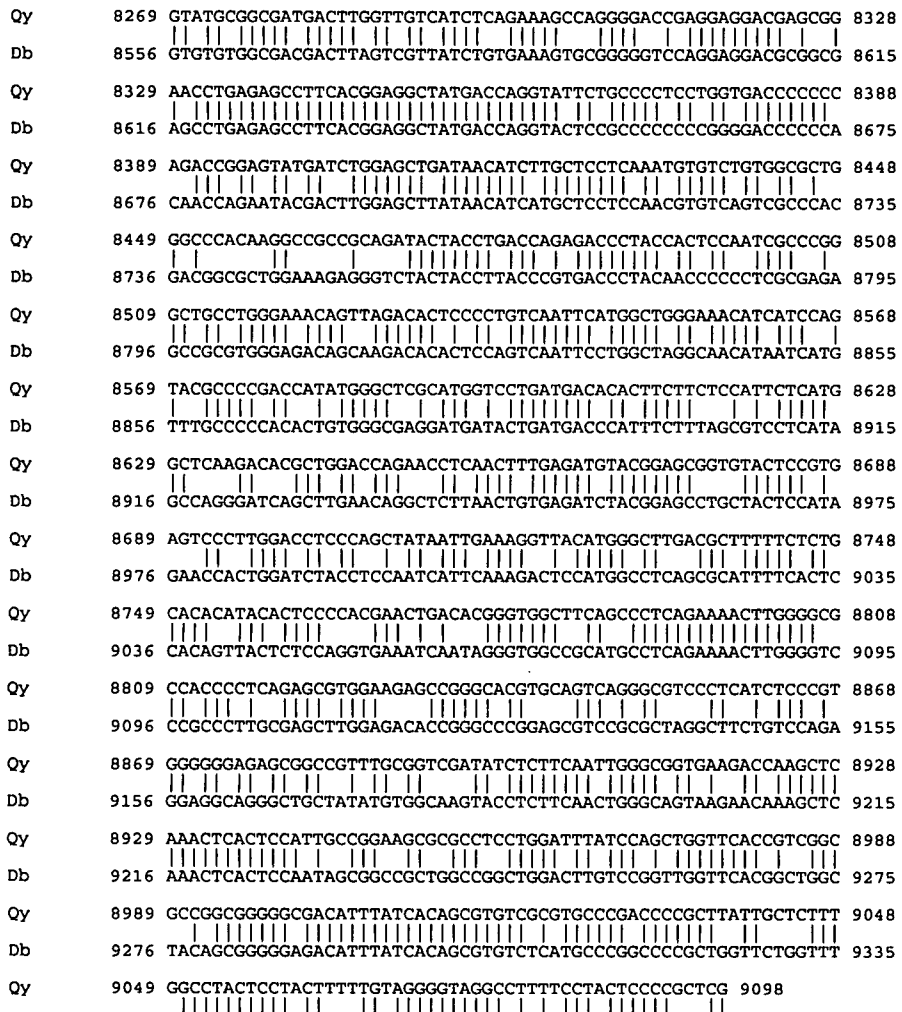
Qy       1621 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1962 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 2021

Qy       1681 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```



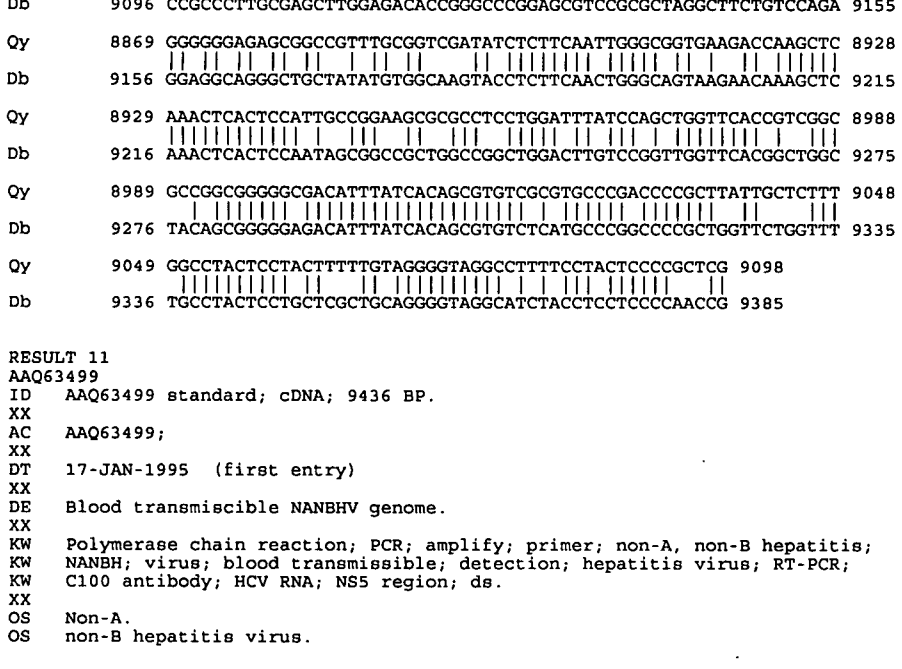


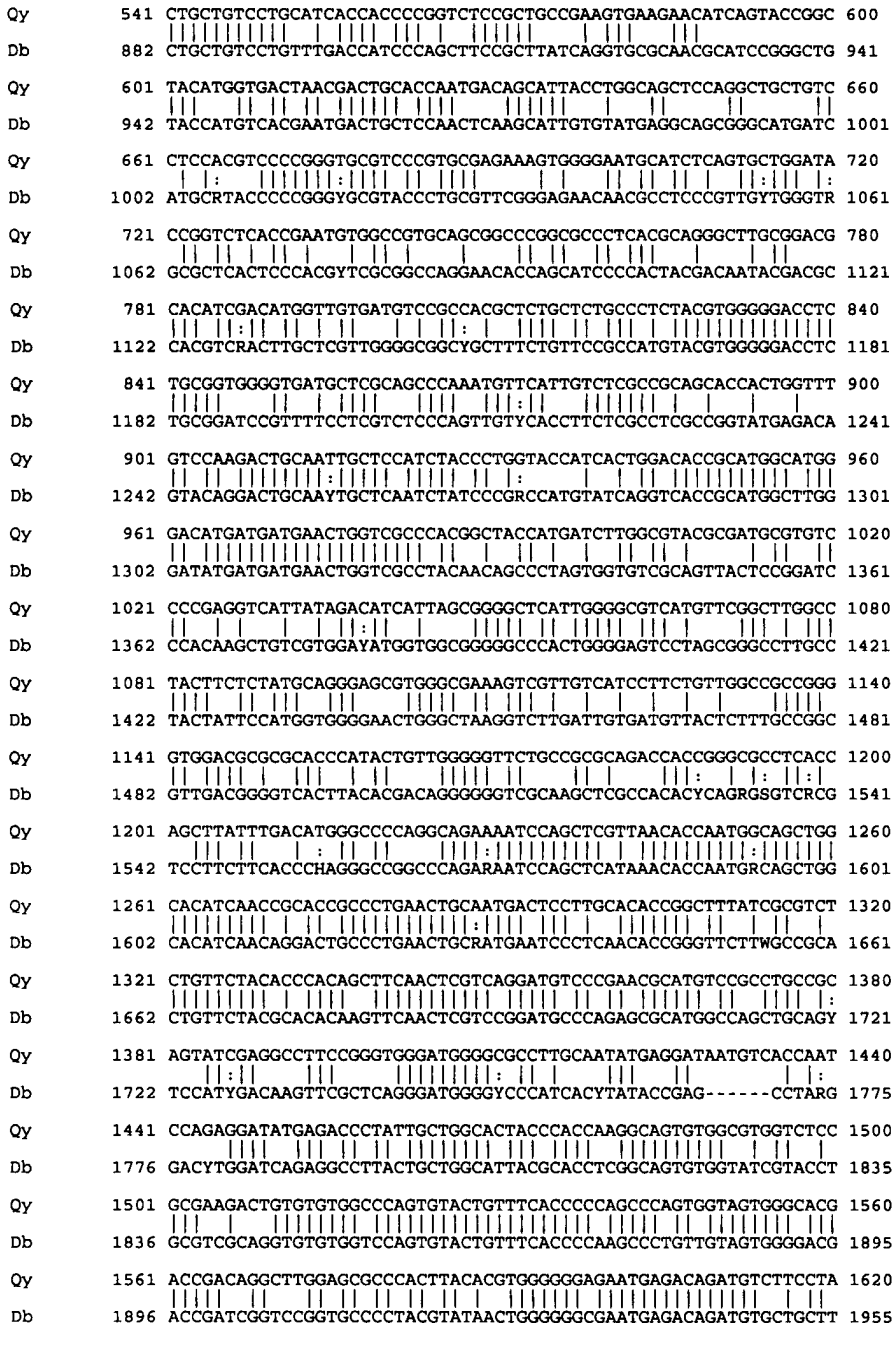

Db 6402 GGGGTCTGGCGAGGAGACGGCATTATGCACACTCGCTGCCACTGTGGAGCTGAGATCACT 6461
Qy 6121 GGCAATGTCCGCTTGGGCTCCATGAGAATCACGGGGCCTAAGACCTGCATGAATATCTGG 6180
Db 6462 GGACATGTCAAAAACGGGACGATGAGGATCGTCGGTCCTAGGACCTGCAGGAACATGTGG 6521
Qy 6181 CAGGGGACCTTTCCTATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCGCCA 6240
Db 6522 AGTGGGACGTTCCCCATTAACGCCTACACCACGGGCCCCTGTACTCCCCTTCCTGCGCCG 6581
Qy 6241 AACTTTAAGGTCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6300
Db 6582 AACTATAAGTTCGCGCTGTGGAGGGTGTCTGCAGAGGAATACGTGGAGATAAGGCGGGTG 6641
Qy 6301 GGGTCATACCACTACATAACAGGACTCACCACTGATAACTTGAAAGTCCCCTGCCAACTA 6360
Db 6642 GGGGACTTCCACTACGTATCGGGTATGACTACTGACAATCTTAAATGCCCGTGCCAGATC 6701
Qy 6361 CCCTCTCCCGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCACA 6420
Db 6702 CCATCGCCCGAATTTTTCACAGAATTGGACGGGGTGCGCCTACACAGGTTTGCGCCCCCT 6761
Qy 6421 CCGAAGCCGTTTTTCCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6480
Db 6762 TGCAAGCCCTTGCTGCGGGAGGAGGTATCATTCAGAGTAGGACTCCACGAGTACCCGGTG 6821
Qy 6481 GGGTCCCAGCTTCCTTGCGACCCTGAACCCGACACAGACGTATTGATGTCCATGCTAACA 6540
Db 6822 GGGTCGCAATTACCTTGCGAGCCCGAACCGGACGTAGCCGTGTTGACGTCCATGCTCACT 6881
Qy 6541 GATCCATCTCATATCACGGCGGAGACTGCAGCGCGGCGTTTAGCGCGGGGGTCACCCCCA 6600
Db 6882 GATCCCTCCCATATAACAGCAGAGGCGGCCGGGAGAAGGTTGGCGAGAGGGTCACCCCCT 6941
Qy 6601 TCCGAGGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 6660
Db 6942 TCTATGGCCAGCTCCTCGGCTAGCCAGCTGTCCGCTCCATCTCTCAAGGCAACTTGCACC 7001
Qy 6661 ACCCACGGCAAAGCCTATGATGTGGACATGGTGGATGCTAACCT----------GTTC 6708
Db 7002 GCCAACCATGACTCCCCTGACGCCGAGCTCATAGAGGCTAACCTCCTGTGGAGGCAGGAG 7061
Qy 6709 ATGGGGGGCGATGTGACTCGGATAGAGTCTGGGTCCAAAGTGGTCGTTCTGGACTCTCTC 6768
Db 7062 ATGGGCGGCAACATCACCAGGGTTGAGTCAGAGAACAAAGTGGTGATTCTGGACTCCTTC 7121
Qy 6769 GACCCAATGGTCGAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCAGAATACATGCTC 6828
Db 7122 GATCCGCTTGTGGCAGAGGAGGATGAGCGGGAGGTCTCCGTACCTGCAGAAATTCTGCGG 7181
Qy 6829 CCCAAGAAGAGGTTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCG 6888
Db 7182 AAGTCTCGGAGATTCGCCCGGGCCCTGCCCGTCTGGGCGCGGCCGGACTACAACCCCCCG 7241
Qy 6889 CTTGTGGAATCGTGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTC 6948
Db 7242 CTAGTAGAGACGTGGAAAAAGCCTGACTACGAACCACCTGTGGTCCATGGCTGCCCGCTA 7301
Qy 6949 CCTCCTCCTAGGAAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGT 7008
Db 7302 CCACCTCCACGGTCCCCTCCTGTGCCTCCGCCTCGGAAAAAGCGTACGGTGGTCCTCACC 7361
Qy 7009 GAGGACTCCATAGGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCC 7068
Db 7362 GAATCAACCCTATCTACTGCCTTGGCCGAGCTTGCCACCAAAAGTTTTGGCAGCTCCTCA 7421
Qy 7069 CCAAGCGGCGATTCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACG 7128
Db 7422 ACTTCCGGCATTACGGGCGACAATACGACAACATCCTCTGAGCCCGCC---------- 7469
Qy 7129 CCTCCTGATGAGTTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCTCGAG 7188
Db 7470 CCTTCTGGCTGCCCCCCCGACTCCGACGTTGAGTCCTATTCTTCCATGCCCCCCCTGGAG 7529
Qy 7189 GGGGAGCTTGGAGATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCAG 7248
Db 7530 GGGGAGCCTGGGGATCCGGA---------- 7549
Qy 7249 GGGGGGGTGGCAGCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGAC 7308
Db 7550 ----------TCTCAGCGACGGGTCATGGTCGACGGTCAGTAGTGGGGCCGACACG 7595
Qy 7309 GACTCCGTCGTGTGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGT 7368
Db 7596 GAAGATGTCGTGTGCTGCTCAATGTCTTATTCCTGGACAGGCGCACTCGTCACCCCGTGC 7655
Qy 7369 AGTCCCGAAGAGGAGAAGTTACCGATTAACCCCTTGAGCAACTCCCTGTTGCGATATCAC 7428
Db 7656 GCTGCGGAAGAACAAAAACTGCCCATCAACGCACTGAGCAACTCGTTGCTACGCCATCAC 7715
Qy 7429 AACAAGGTGTACTGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTT 7488
Db 7716 AATCTGGTGTATTCCACCACTTCACGCAGTGCTTGCCAAAGGCAGAAGAAAGTCACATTT 7775
Qy 7489 GATAGGATGCAAGTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCG 7548
Db 7776 GACAGACTGCAAGTTCTGGACAGCCATTACCAGGACGTGCTCAAGGAGGTCAAAGCAGCG 7835
Qy 7549 GCCTCCAAGGTCACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCC 7608
Db 7836 GCGTCAAAAGTGAAGGCTAACTTGCTATCCGTAGAGGAAGCTTGCAGCCTGACGCCCCCA 7895
Qy 7609 CATTCTGCAAGATCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGG 7668
Db 7896 CATTCAGCCAAATCCAAGTTTGGCTATGGGGCAAAAGACGTCCGTTGCCATGCCAGAAAG 7955
Qy 7669 GCCGTTAACCACATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATT 7728
Db 7956 GCCGTAGCCCACATCAACTCCGTGTGGAAAGACCTTCTGGAAGACAGTGTAACACCAATA 8015
Qy 7729 CCCACAACCATTATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAG 7788
Db 8016 GACACTACCATCATGGCCAAGAACGAGGTTTTCTGCGTTCAGCCTGAGAAGGGGGGTCGT 8075
Qy 7789 AAAGCAGCTCGCCTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCC 7848
Db 8076 AAGCCAGCTCGTCTCATCGTGTTCCCCGACCTGGGCGTGCGCGTGTGCGAGAAGATGGCC 8135
Qy 7849 CTTTATGACATTACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAG 7908
Db 8136 CTGTACGACGTGGTTAGCAAGCTCCCCCTGGCCGTGATGGGAAGCTCCTACGGATTCCAA 8195
Qy 7909 TATTCCCCCGCTCAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCT 7968
Db 8196 TACTCACCAGGACAGCGGGTTGAATTCCTCGTGCAAGCGTGGAAGTCCAAGAAGACCCCG 8255
Qy 7969 ATGGGTTTTTCGTATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGG 8028
Db 8256 ATGGGGTTCTCGTATGATACCCGCTGTTTTGACTCCACAGTCACTGAGAGCGACATCCGT 8315
Qy 8029 ACTGAGGAGTCCATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACAC 8088
Db 8316 ACGGAGGAGGCAATTTACCAATGTTGTGACCTGGACCCCCAAGCCCGCGTGGCCATCAAG 8375
Qy 8089 TCGCTAACTGAGAGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGC 8148
Db 8376 TCCCTCACTGAGAGGCTTTATGTTGGGGGCCCTCTTACCAATTCAAGGGGGGAAAACTGC 8435
Qy 8149 GGGTACAGGCGTTGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACA 8208
Db 8436 GGCTACCGCAGGTGCCGCGCGAGCGGCGTACTGACAACTAGCTGTGGTAACACCCTCACT 8495
Qy 8209 TGCTACGTGAAAGCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTG 8268
Db 8496 TGCTACATCAAGGCCCGGGCAGCCTGTCGAGCCGCAGGGCTCCAGGACTGCACCATGCTC 8555

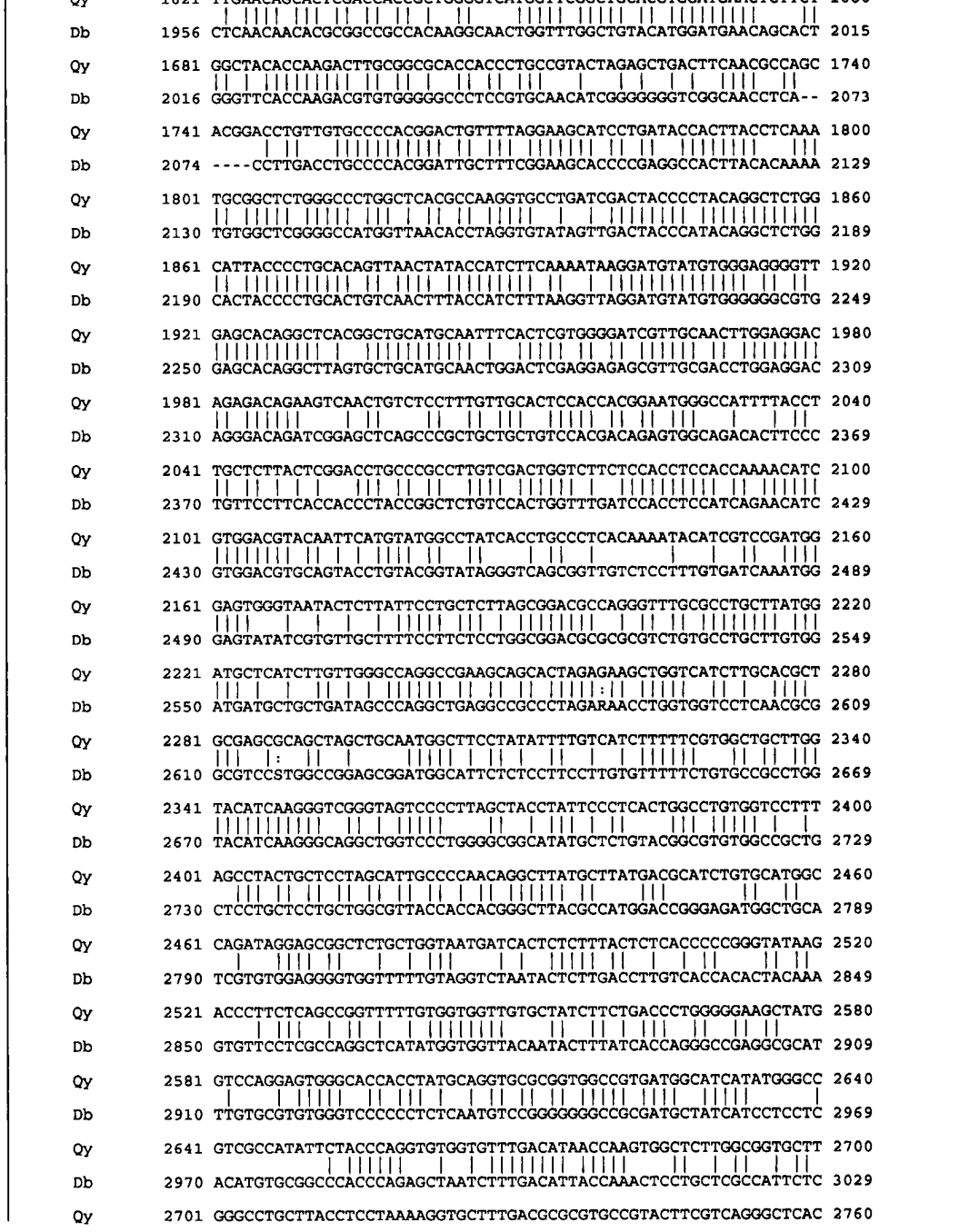


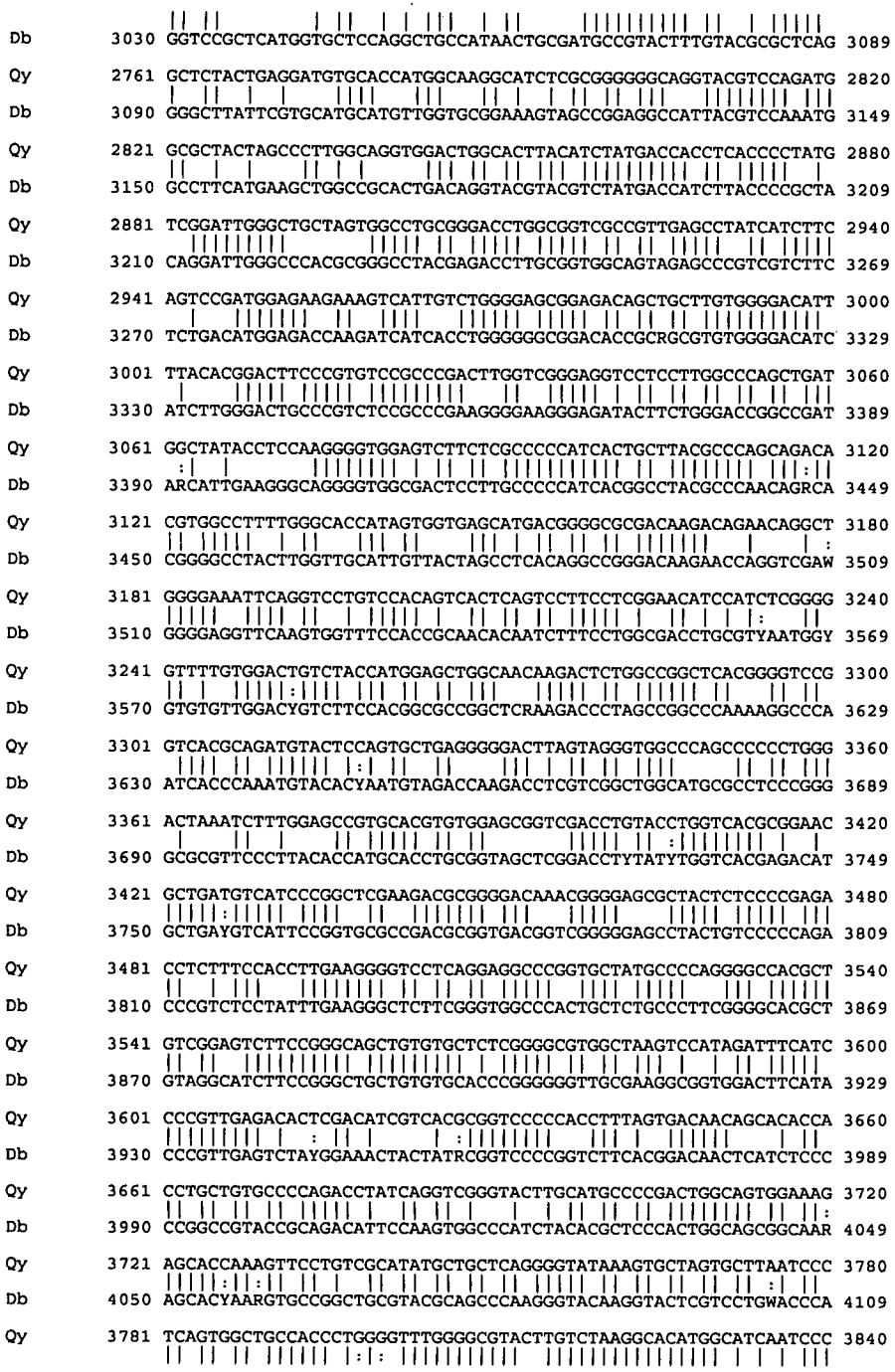
```
RESULT 11
AAQ63499
ID   AAQ63499 standard; cDNA; 9436 BP.
XX
AC   AAQ63499;
XX
DT   17-JAN-1995  (first entry)
XX
DE   Blood transmiscible NANBHV genome.
XX
KW   Polymerase chain reaction; PCR; amplify; primer; non-A, non-B hepatitis;
KW   NANBH; virus; blood transmissible; detection; hepatitis virus; RT-PCR;
KW   C100 antibody; HCV RNA; NS5 region; ds.
XX
OS   Non-A.
OS   non-B hepatitis virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             342. .9374
FT                   /*tag=  a
XX
PN   JP06105690-A.
XX
PD   19-APR-1994.
XX
PF   10-MAR-1992;   92JP-00051885.
XX
PR   10-MAR-1992;   92JP-00051885.
XX
PA   (KAEN/) KAENNO K.
XX
DR   WPI; 1994-163130/20.
DR   P-PSDB; AAR53417.
XX
PT   Blood-transmissible non-A non-B hepatitis virus DNA - used for detection
PT   of hepatitis virus.
XX
PS   Claim 1; Page 8-20; 22pp; Japanese.
XX
CC   This sequence represents the genome of a blood transmissible non-A, non-B
CC   hepatitis (NANBH) virus. This sequence was isolated using the primers
CC   given in AAQ63500-35. The amplified fragments are used in the detection
CC   of hepatitis virus. This target DNA was isolated from serum of
CC   chronically infected NANBH patients who were C100 antibody-positive and
CC   HCV RNA (NS5 region) positive. Reverse transcription-PCR and PCR were
CC   performed on cDNA and the total human NANBH DNA was constructed from 23
CC   clones
XX
SQ   Sequence 9436 BP; 1876 A; 2840 C; 2655 G; 1974 T; 0 U; 91 Other;

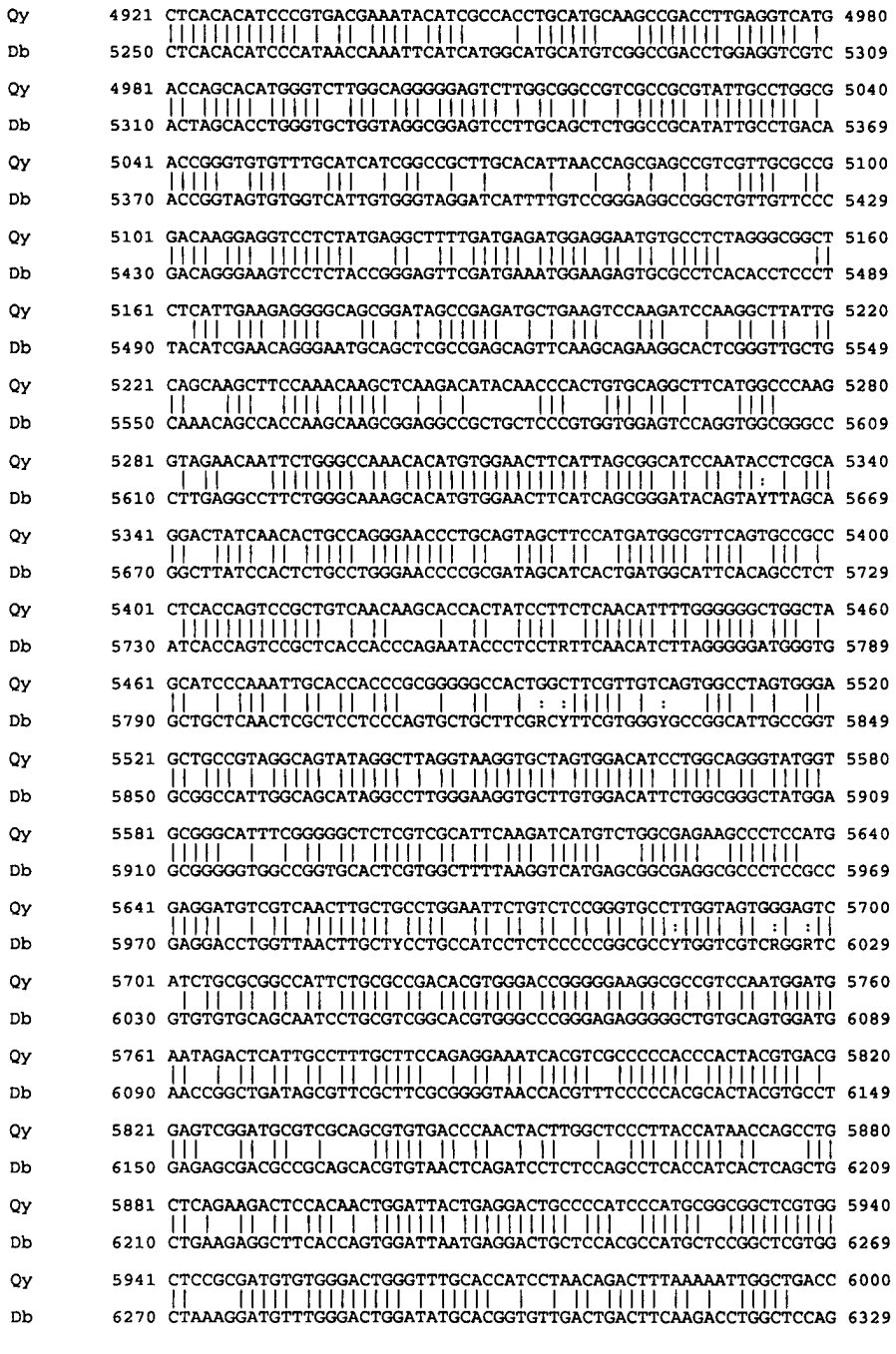
  Query Match             45.7%;  Score 4160.2;  DB 2;  Length 9436;
  Best Local Similarity   66.2%;  Pred. No. 0;
  Matches 6029;  Conservative   65;  Mismatches 2924;  Indels   93;  Gaps    6;
```

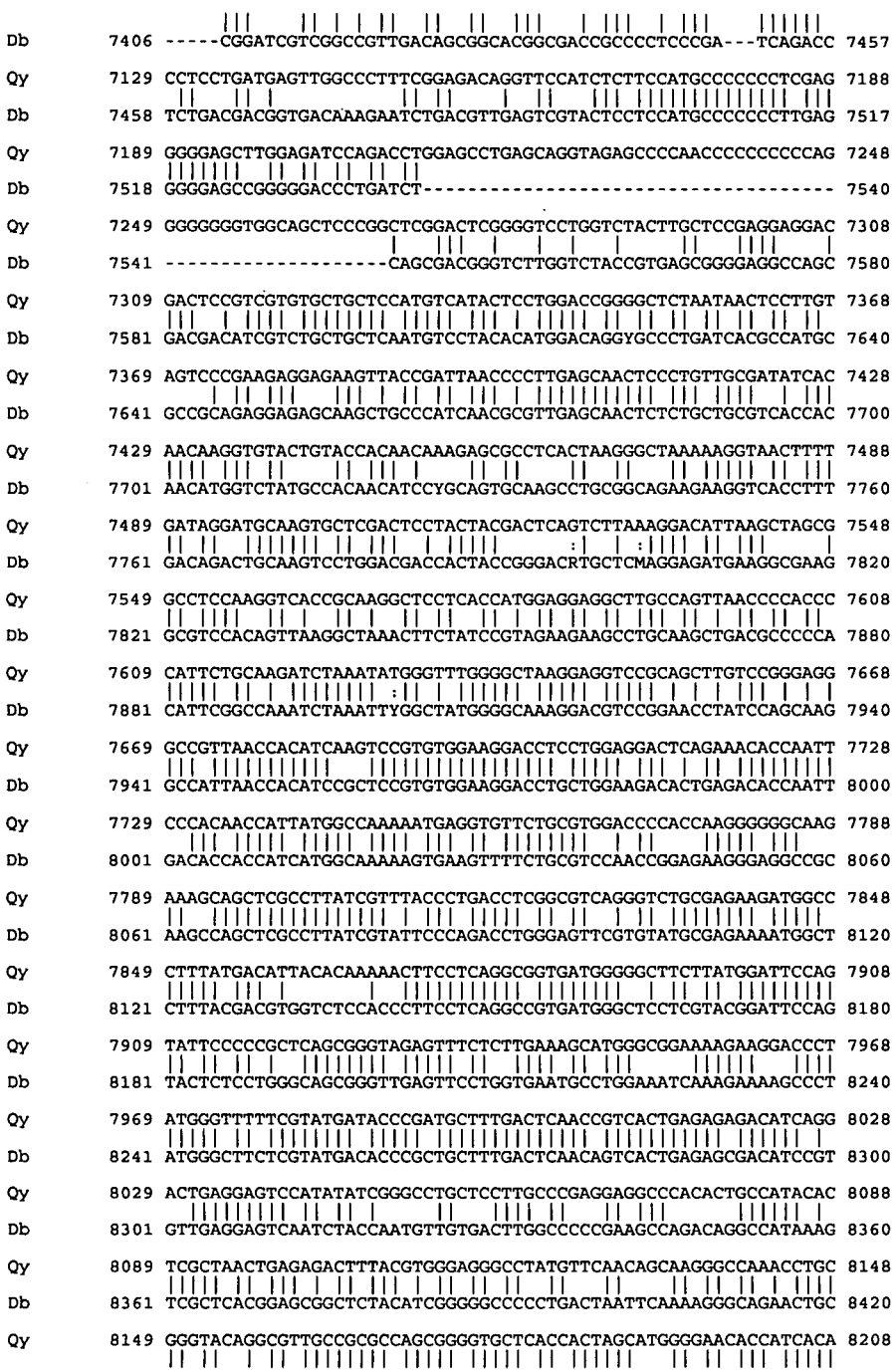


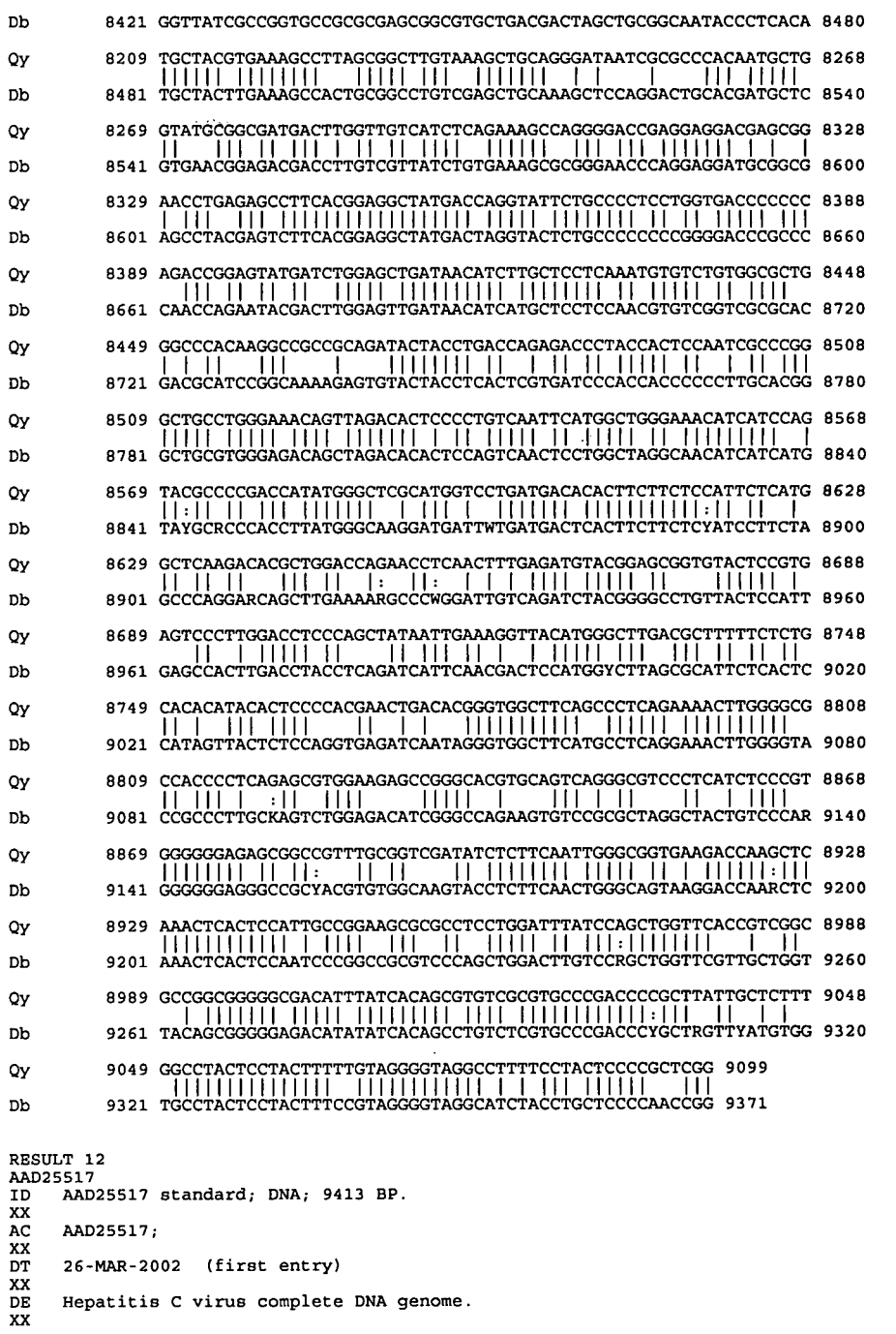





Db 3030 GGTCCGCTCATGGTGCTCCAGGCTGCCATAACTGCGATGCCGTACTTTGTACGCGCTCAG 3089
Qy 2761 GCTCTACTGAGGATGTGCACCATGGCAAGGCATCTCGCGGGGGGCAGGTACGTCCAGATG 2820
Db 3090 GGGCTTATTCGTGCATGCATGTTGGTGCGGAAAGTAGCCGGAGGCCATTACGTCCAAATG 3149
Qy 2821 GCGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 2880
Db 3150 GCCTTCATGAAGCTGGCCGCACTGACAGGTACGTACGTCTATGACCATCTTACCCCGCTA 3209
Qy 2881 TCGGATTGGGCTGCTAGTGGCCTGCGGGACCTGGCGGTCGCCGTTGAGCCTATCATCTTC 2940
Db 3210 CAGGATTGGGCCCACGCGGGCCTACGAGACCTTGCGGTGGCAGTAGAGCCCGTCGTCTTC 3269
Qy 2941 AGTCCGATGGAGAAGAAAGTCATTGTCTGGGGAGCGGAGACAGCTGCTTGTGGGGACATT 3000
Db 3270 TCTGACATGGAGACCAAGATCATCACCTGGGGGGCGGACACCGCRGCGTGTGGGGACATC 3329
Qy 3001 TTACACGGACTTCCCGTGTCCGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3060
Db 3330 ATCTTGGGACTGCCCGTCTCCGCCCGAAGGGGAAGGGAGATACTTCTGGGACCGGCCGAT 3389
Qy 3061 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3120
Db 3390 ARCATTGAAGGGCAGGGGTGGCGACTCCTTGCCCCCATCACGGCCTACGCCCAACAGRCA 3449
Qy 3121 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCT 3180
Db 3450 CGGGGCCTACTTGGTTGCATTGTTACTAGCCTCACAGGCCGGGACAAGAACCAGGTCGAW 3509
Qy 3181 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3240
Db 3510 GGGGAGGTTCAAGTGGTTTCCACCGCAACACAATCTTTCCTGGCGACCTGCGTYAATGGY 3569
Qy 3241 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3300
Db 3570 GTGTGTTGGACYGTCTTCCACGGCGCCGGCTCRAAGACCCTAGCCGGCCCAAAAGGCCCA 3629
Qy 3301 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGG 3360
Db 3630 ATCACCCAAATGTACACYAATGTAGACCAAGACCTCGTCGGCTGGCATGCGCCTCCCGGG 3689
Qy 3361 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3420
Db 3690 GCGCGTTCCCTTACACCATGCACCTGCGGTAGCTCGGACCTYTATYTGGTCACGAGACAT 3749
Qy 3421 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3480
Db 3750 GCTGAYGTCATTCCGGTGCGCCGACGCGGTGACGGTCGGGGGAGCCTACTGTCCCCCAGA 3809
Qy 3481 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3540
Db 3810 CCCGTCTCCTATTTGAAGGGCTCTTCGGGTGGCCCACTGCTCTGCCCTTCGGGGCACGCT 3869
Qy 3541 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3600
Db 3870 GTAGGCATCTTCCGGGCTGCTGTGTGCACCCGGGGGGTTGCGAAGGCGGTGGACTTCATA 3929
Qy 3601 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 3660
Db 3930 CCCGTTGAGTCTAYGGAAACTACTATRCGGTCCCCGGTCTTCACGGACAACTCATCTCCC 3989
Qy 3661 CCTGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 3720
Db 3990 CCGGCCGTACCGCAGACATTCCAAGTGGCCCATCTACACGCTCCCACTGGCAGCGGCAAR 4049
Qy 3721 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 3780
Db 4050 AGCACYAARGTGCCGGCTGCGTACGCAGCCCAAGGGTACAAGGTACTCGTCCTGWACCCA 4109
Qy 3781 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 3840
Db 4110 TCTGTTGCCGCCACCTTRGRTTTTGGGGCGTATATGTCTAAGGCACATGGTACCGACCCC 4169
Qy 3841 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 3900
Db 4170 AACATCAGRACTGGGGTAAGGACCATTACCACGGGCGCCCCCATCACGTACTCCACCTAT 4229
Qy 3901 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 3960
Db 4230 GGTAAGTTCCTTGCCGACGGTGGATGTTCTGGGGGCGCCTATGATATCATAATATGTGAT 4289
Qy 3961 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4020
Db 4290 GAGTGCCACTCAACTGACTCGACTACCATCCTGGGCATCGGCACAGTCCTGGACCAAGCG 4349
Qy 4021 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4080
Db 4350 GAGACGGCTGGAGCGCGGCTCGTCGTGCTCGCCACCGCTACGCCTCCGGGATCGGTTACC 4409
Qy 4081 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4140
Db 4410 GTGCCACACCCAAATATTGAGGAGGTGGCTCTGTCCAACACTGGAGAGATCCCCTTCTAT 4469
Qy 4141 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4200
Db 4470 GGCAAGGCCATCCCCCTCGAGGCCATCAAGGGGGGAAGGCATCTCATTTTCTGCCACTCC 4529
Qy 4201 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4260
Db 4530 AAGAAGAAGTGCGATGAGCTCGCCGCAAAGCTGTCAGGCCTCGGAATCAATGCCGTAGCG 4589
Qy 4261 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4320
Db 4590 TATTACCGGGGTCTTGACGTGTCCGTCATACCGACTAGCGGAGATGTCGTTATCGTGGCA 4649
Qy 4321 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4380
Db 4650 ACAGACGCTCTAATGACGGGCTACACCGGTGACTTTGACTCGGTGATCGACTGCRACACA 4709
Qy 4381 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4440
Db 4710 TGTGTCACCCAGACRGTCGATTTCAGCTTGGATCCCACCTTTACCATTGAGACGACGACC 4769
Qy 4441 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4500
Db 4770 GTGCCCCAAGACGCGGTGTCGCGCTCACAACGGCGAGGCAGRACTGGCAGGGGTAGGGGA 4829
Qy 4501 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4560
Db 4830 GGCATCTACAGGTTTGTGACTCCGGGAGAACGGCCCTCGGGCATGTTCGATTCTTCSGTC 4889
Qy 4561 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4620
Db 4890 CTGTGTGAGTGCTATGACGCGGGCTGTGCTTGGTATGAGCTCACGCCCGCTGAGACTACA 4949
Qy 4621 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 4680
Db 4950 GTCAGGYTGCGGGCTTACCTGAATACACCRGGGTTGCCCGTCTGCCAGGACCATCTGGAG 5009
Qy 4681 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 4740
Db 5010 TTCTGGGAAAGCGTCTTCACAGGCCTCACCCACATAGATGCCCACTTCCTGTCCCAAACC 5069
Qy 4741 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 4800
Db 5070 AAGCAGGCAGGAGACAACTTCCCTTACCTGGTAGCATACCAAGCCACGGTGTGCGCCAGG 5129
Qy 4801 GCCAAAGCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 4860
Db 5130 GCTCAGGCTCCACCTCCATCGTGGGATCAAATGTGGAAGTGTCTCATACGGCTTAAACCT 5189
Qy 4861 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 4920
Db 5190 ACGCTGCACGGGCCAACACCCCTGCTGTATAGGCTAGGAGCCGTTCAAAATGAGATCACC 5249


Qy 4921 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 4980
Db 5250 CTCACACATCCCATAACCAAATTCATCATGGCATGCATGTCGGCCGACCTGGAGGTCGTC 5309
Qy 4981 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5040
Db 5310 ACTAGCACCTGGGTGCTGGTAGGCGGAGTCCTTGCAGCTCTGGCCGCATATTGCCTGACA 5369
Qy 5041 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5100
Db 5370 ACCGGTAGTGTGGTCATTGTGGGTAGGATCATTTTGTCCGGGAGGCCGGCTGTTGTTCCC 5429
Qy 5101 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5160
Db 5430 GACAGGGAAGTCCTCTACCGGGAGTTCGATGAAATGGAAGAGTGCGCCTCACACCTCCCT 5489
Qy 5161 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5220
Db 5490 TACATCGAACAGGGAATGCAGCTCGCCGAGCAGTTCAAGCAGAAGGCACTCGGGTTGCTG 5549
Qy 5221 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5280
Db 5550 CAAACAGCCACCAAGCAAGCGGAGGCCGCTGCTCCCGTGGTGGAGTCCAGGTGGCGGGCC 5609
Qy 5281 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5340
Db 5610 CTTGAGGCCTTCTGGGCAAAGCACATGTGGAACTTCATCAGCGGGATACAGTAYTTAGCA 5669
Qy 5341 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5400
Db 5670 GGCTTATCCACTCTGCCTGGGAACCCCGCGATAGCATCACTGATGGCATTCACAGCCTCT 5729
Qy 5401 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5460
Db 5730 ATCACCAGTCCGCTCACCACCCAGAATACCCTCCTRTTCAACATCTTAGGGGGATGGGTG 5789
Qy 5461 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5520
Db 5790 GCTGCTCAACTCGCTCCTCCCAGTGCTGCTTCGRCYTTCGTGGGYGCCGGCATTGCCGGT 5849
Qy 5521 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5580
Db 5850 GCGGCCATTGGCAGCATAGGCCTTGGGAAGGTGCTTGTGGACATTCTGGCGGGCTATGGA 5909
Qy 5581 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5640
Db 5910 GCGGGGGTGGCCGGTGCACTCGTGGCTTTTAAGGTCATGAGCGGCGAGGCGCCCTCCGCC 5969
Qy 5641 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 5700
Db 5970 GAGGACCTGGTTAACTTGCTYCCTGCCATCCTCTCCCCCGGCGCCYTGGTCGTCRGGRTC 6029
Qy 5701 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 5760
Db 6030 GTGTGTGCAGCAATCCTGCGTCGGCACGTGGGCCCGGGAGAGGGGGCTGTGCAGTGGATG 6089
Qy 5761 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 5820
Db 6090 AACCGGCTGATAGCGTTCGCTTCGCGGGGTAACCACGTTTCCCCCACGCACTACGTGCCT 6149
Qy 5821 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 5880
Db 6150 GAGAGCGACGCCGCAGCACGTGTAACTCAGATCCTCTCCAGCCTCACCATCACTCAGCTG 6209
Qy 5881 CTCAGAAGACTCCACAACTGGATTACTGAGGACTGCCCCATCCCATGCGGCGGCTCGTGG 5940
Db 6210 CTGAAGAGGCTTCACCAGTGGATTAATGAGGACTGCTCCACGCCATGCTCCGGCTCGTGG 6269
Qy 5941 CTCCGCGATGTGTGGGACTGGGTTTGCACCATCCTAACAGACTTTAAAAATTGGCTGACC 6000
Db 6270 CTAAAGGATGTTTGGGACTGGATATGCACGGTGTTGACTGACTTCAAGACCTGGCTCCAG 6329
Qy 6001 TCCAAATTATTCCCAAAGATGCCCGGCCTCCCCTTTGTCTCCTGTCAAAAGGGGTACAAG 6060
Db 6330 TCCAAGCTCCTGCCGAAATTGCCGGGAGTCCCTTTCTTCTCATGCCAACGCGGGTACAAG 6389
Qy 6061 GGCGTGTGGGCCGGCACTGGCATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6120
Db 6390 GGAGTCTGGCGGGGGGACGGCATCATGCAAACCACCTGCCCATGTGGAGCACAAATTACC 6449
Qy 6121 GGCAATGTCCGCTTGGGCTCCATGAGAATCACGGGGCCTAAGACCTGCATGAATATCTGG 6180
Db 6450 GGACATGTCAAAAACGGTTCCATGAGGATCGTTGGGCCTAAAACCTGCAGCAACACGTGG 6509
Qy 6181 CAGGGGACCTTTCCTATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCGCCA 6240
Db 6510 CACGGAACGTTCCCCATCAACGCGTACACCACAGGCCCCTGCACACCCTCCCCGGCGCCG 6569
Qy 6241 AACTTTAAGGTCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6300
Db 6570 AACTATTCCAGGGCGCTGTGGCGGGTGGCTGCTGAAGAGTACGTGGAGATTACGCGGGTG 6629
Qy 6301 GGGTCATACCACTACATAACAGGACTCACCACTGATAACTTGAAAGTCCCCTGCCAACTA 6360
Db 6630 GGGGACTTCCACTACGTGACGGGTATGACCACCGACAACGTAAAATGCCCGTGCCAGGTC 6689
Qy 6361 CCCTCTCCCGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCACA 6420
Db 6690 CCGGCCCCCGAATTCTTYACCGAATTGGACGGGGTGCGGTTGCACAGGTACGCTCCGGCG 6749
Qy 6421 CCGAAGCCGTTTTTCCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6480
Db 6750 TGCAGACCTCTCCTACGGGWGGAKGTCACATTCCAGGTCGGGCTCAACCAATACCTGGTT 6809
Qy 6481 GGGTCCCAGCTTCCTTGCGACCCTGAACCCGACACAGACGTATTGATGTCCATGCTAACA 6540
Db 6810 GGGTCACAGCTCCCATGCGAGCCTGARCCGGATGTGGCAGTGCTCACTTCCATGCTCACC 6869
Qy 6541 GATCCATCTCATATCACGGCGGAGACTGCAGCGCGGCGTTTAGCGCGGGGGTCACCCCCA 6600
Db 6870 GACCCCTCCCACATTACAGCAGAGACGGCTAAACGTAGGCTGGCCAGGGGGTCTCCCCCC 6929
Qy 6601 TCCGAGGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 6660
Db 6930 TCCTTGGCCAGCTCTTCAGCTAGCCAATTGTCTGCGCCTTCCTTGAAGGCAACATGCACT 6989
Qy 6661 ACCCACGGCAAAGCCTATGATGTGGACATGGTGGATGCTAACCT----------GTTC 6708
Db 6990 ACCCACCATGACTCCCCGGACGCTGACCTCATCGAGGCCAACCTCCTGTGGCGGCAGGAG 7049
Qy 6709 ATGGGGGGCGATGTGACTCGGATAGAGTCTGGGTCCAAAGTGGTCGTTCTGGACTCTCTC 6768
Db 7050 ATGGGCGGAAACATCACCCGTGTGGAGTCAGAGAATAAGGTGGTAATTTTGGACTCTTTC 7109
Qy 6769 GACCCAATGGTCGAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCAGAATACATGCTC 6828
Db 7110 GACCCGCTTCGAGCGGAAGAGGATGAGAGGGAAGTATCCGTTGCAGCAGAGATCCTGCGA 7169
Qy 6829 CCCAAGAAGAGGTTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCG 6888
Db 7170 AAATCCAAGAAGTTCCCCCCCGCGTTGCCCATATGGGCACGCCCGGATTACAACCCTCCA 7229
Qy 6889 CTTGTGGAATCGTGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTC 6948
Db 7230 CTGTTAGAGTCCTGGAAAAGTCCGGACTACGTCCCTCCGGCGGTGCATGGGTGCCCATTG 7289
Qy 6949 CCTCCTCCTAGGAAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGT 7008
Db 7290 CCGCCTACCACGGGCCCTCCAATACCGCCTCCACGGAAAAAGAGGACGGTTGTTCTGACA 7349
Qy 7009 GAGGACTCCATAGGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCC 7068
Db 7350 GAGTCCACCGTGTCTTCTGCCTTGGCGGAGCTGGCTACTAAGACTTTCGGCAGCTC---- 7405
Qy 7069 CCAAGCGGCGATTCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACG 7128





RESULT 12
AAD25517
ID AAD25517 standard; DNA; 9413 BP.
XX
AC AAD25517;
XX
DT 26-MAR-2002 (first entry)
XX
DE Hepatitis C virus complete DNA genome.
XX

KW HCV; hepatitis C virus; cytostatic; cancer; immunosuppressive; virucide;
KW antibacterial; fungicide; protozoacide; antirheumatic; antiinflammatory;
KW antiarthritic; rheumatoid arthritis; neuroprotective; multiple sclerosis;
KW immune response; vasotropic; vaccine; gene therapy; autoimmune disease;
KW vasculitis; ds.
XX
OS Hepatitis C virus.
XX
PN WO200176643-A1.
XX
PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001WO-US011372.
XX
PR 07-APR-2000; 2000US-0195680P.
XX
PA (BAYU) BAYLOR COLLEGE MEDICINE.
XX
PI Orson FM, Kinsey BM, Bhogal BS;
XX
DR WPI; 2002-066308/09.
XX
PT Composition for oral delivery of vaccines, comprises expression vector
PT containing antigenic genomic sequence, bound to aggregated protein-
PT polycationic polymer conjugate or suspension.
XX
PS Disclosure; Page 84-87; 145pp; English.
XX
CC The invention relates to a composition comprising an expression vector
CC bound to an aggregated protein-polycationic polymer conjugate or
CC suspension. The expression vector contains a promoter polynucleotide
CC sequence operatively linked to a polynucleotide sequence encoding an
CC antigen which is a fragment of a gene or genome associated with an
CC infectious disease, cancer and autoimmune disease such as rheumatoid
CC arthritis, vasculitis, and multiple sclerosis, pathogenic genomes
CC consisting of bacterium, fungus, protozoa and virus such as human
CC immunodeficiency virus (HIV), herpes simplex virus (HSV), hepatitis C
CC virus (HCV), influenza and respiratory syncytial virus (RSV), and
CC optionally comprising a nucleotide sequence encoding a cytokine (or a
CC cytokine expression vector), is useful for inducing an immune response
CC (systemic and/or mucosal) in an organism. The cytokine expression vector
CC contains a sequence for granulocyte macrophage-colony stimulating factor
CC (GM-CSF) or interleukin-12 (IL-12). The polynucleotide sequences encoding
CC the antigen and the cytokine are under transcriptional control of same or
CC different promoter polynucleotide sequences. The expression vector, as a
CC DNA vaccine is useful for treating a condition in an organism. The
CC present sequence is hepatitis C virus complete DNA genome related to the
CC invention
XX
SQ Sequence 9413 BP; 1886 A; 2816 C; 2694 G; 2017 T; 0 U; 0 Other;

Query Match 45.5%; Score 4141.8; DB 6; Length 9413;
Best Local Similarity 66.6%; Pred. No. 0;
Matches 6066; Conservative 0; Mismatches 2952; Indels 93; Gaps 6;

```
Qy          1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 60
              |||||||||||||||||||||||||||||||||||| | ||||||||||| ||||||||
Db        330 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCCGCCCACAG 389

Qy         61 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 120
              |||||||||| |||||||| || |||||||||| ||||| ||| ||||||||||||||
Db        390 GACGTTAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGG 449

Qy        121 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 180
              |||||||||||||||||||||||| |||||||||| |||||||| || || ||||||
Db        450 GGCCCCAGGTTGGGTGTGCGCGCGACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTGGA 509

Qy        181 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 240
              ||||| || || |||||| || | ||| ||  ||   || |   ||||||   | || ||
Db        510 AGGCGACAACCTATCCCCAAGGCTCGCCGGCCCGAGGGTAGGACCTGGGCTCAGCCCGGG 569

Qy        241 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 300
              ||||| |||||||| || || || |||||  | || ||||||||||||||||||| |||
```

```
Db        570 TACCCTTGGCCCCTCTATGGCAACGAGGGTATGGGGTGGGCAGGATGGCTCCTGTCACCC 629

Qy        301 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 360
              || || || || ||    |||||||||| |||||||| ||||||||| ||  |||||
Db        630 CGTGGCTCTCGGCCTAGTTGGGGCCCCACAGACCCCCGGCGTAGGTCGCGTAATTTGGGT 689

Qy        361 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 420
              |||||||||||||||| || |||||||| |||||||||||||||||| ||  | ||
Db        690 AAGGTCATCGATACCCTTACATGCGGCTTCGCCGACCTCATGGGGTACATTCCGCTTGTC 749

Qy        421 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 480
              |||||||| || || |||| ||||| || || || ||||| ||  | || |||||||||
Db        750 GGCGCCCCCCTAGGGGGCGCTGCCAGGGCCCTGGCACATGGTGTCCGGGTTCTGGAGGAC 809

Qy        481 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 540
              || || || | |||||||||||| | |||||||||| || |||||||| |  | ||
Db        810 GGCGTGAACTATGCAACAGGGAATCTGCCCGGTTGCTCTTTCTCTATCTTCCTCTTAGCT 869

Qy        541 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 600
              ||||||| ||  | |||| ||| |  |||||  ||| |||   ||| |
Db        870 TTGCTGTCTTGTTTGACCATCCCAGCTTCCGCTTACGAGGTGCGCAACGTGTCCGGGATA 929

Qy        601 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 660
              |||   || || ||||||||| ||||    || |||   |   ||     ||     ||
Db        930 TACCATGTCACGAACGACTGCTCCAACTCAAGTATTGTGTATGAGGCAGCGGACATGATC 989

Qy        661 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 720
              | |||  |||||||||||| || ||||      | | | |||   || |  |||||| ||
Db        990 ATGCACACCCCCGGGTGCGTGCCCTGCGTCCGGGAGAGTAATTTCTCCCGTTGCTGGGTA 1049

Qy        721 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 780
              || || | || |   | || |   |   | ||  || |||   |   | ||
Db       1050 GCGCTCACTCCCACGCTCGCGGCCAGGAACAGCAGCATCCCCACCACGACAATACGACGC 1109

Qy        781 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 840
              ||| ||||  || | ||   | | ||  | ||||| || ||  | ||||| ||||| |||
Db       1110 CACGTCGATTTGCTCGTTGGGGCGGCTGCTCTCTGTTCCGCTATGTACGTTGGGGATCTC 1169

Qy        841 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 900
              |||||     ||  | ||||   ||||  ||||||   |||| || |  |   |   |
Db       1170 TGCGGATCCGTTTTTCTCGTCTCCCAGCTGTTCACCTTCTCACCTCGCCGGTATGAGACG 1229

Qy        901 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 960
              || ||||| |||||||||| ||||| || ||   | |  | || |||||||||| |||
Db       1230 GTACAAGATTGCAATTGCTCAATCTATCCCGGCCACGTATCAGGTCACCGCATGGCTTGG 1289

Qy        961 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1020
              || |||||||||||||||| || ||  |  || |  |  |  || |     | ||  ||
Db       1290 GATATGATGATGAACTGGTCACCTACAACGGCCCTAGTGGTATCGCAGCTACTCCGGATC 1349

Qy       1021 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1080
              ||  | | | |  | ||||  |  ||||| || ||||| ||| |   ||| | |||
Db       1350 CCACAAGCCGTCGTGGACATGGTGGCGGGGGCCCACTGGGGTGTCCTAGCGGGCCTTGCC 1409

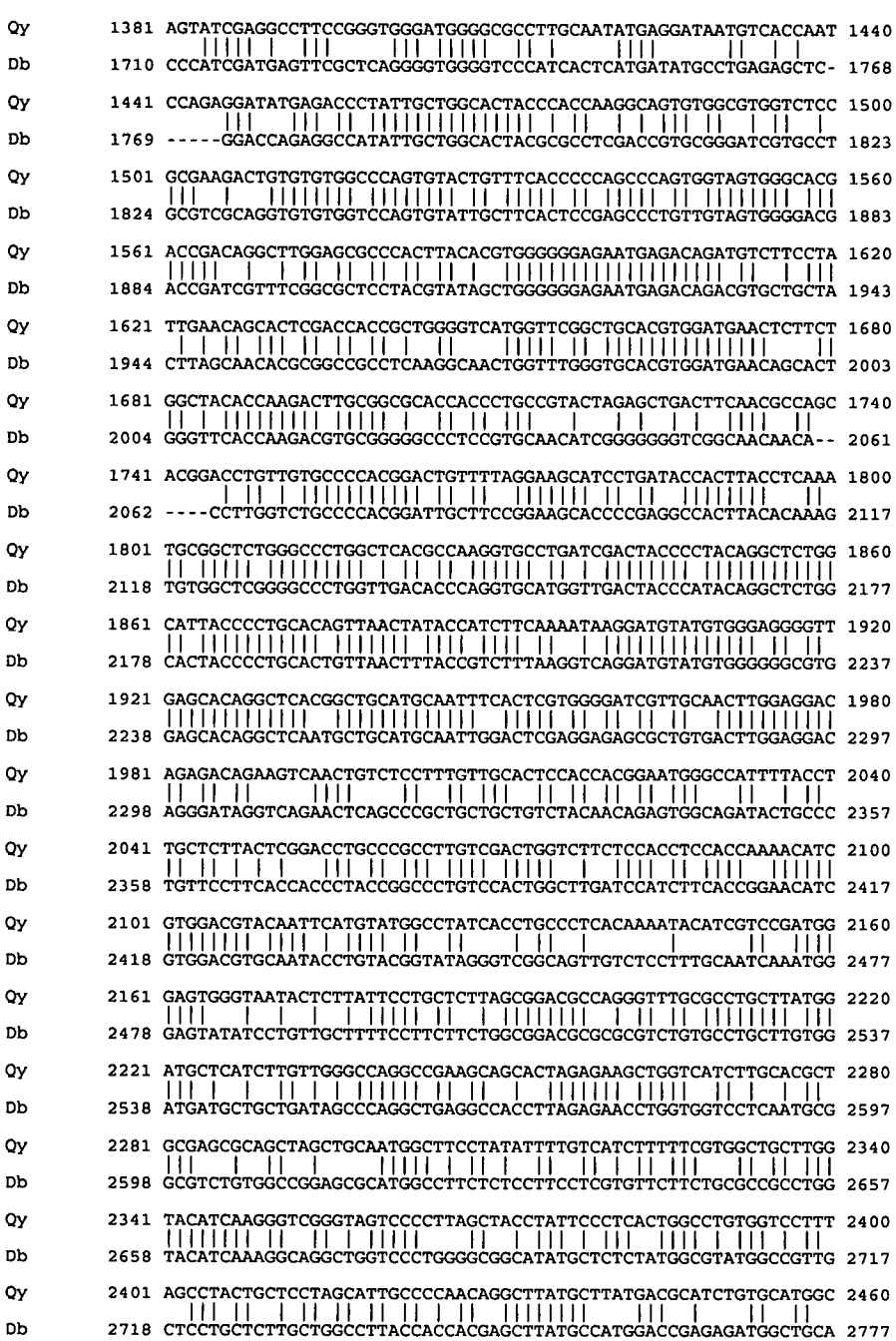
Qy       1081 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1140
              ||||  || ||| |||    ||||| || ||| |  |  | || |    || ||
Db       1410 TACTATTCCATGGTGGGGAACTGGGCTAAGGTCTTGATTGTGATGCTACTCTTTGCTGGC 1469

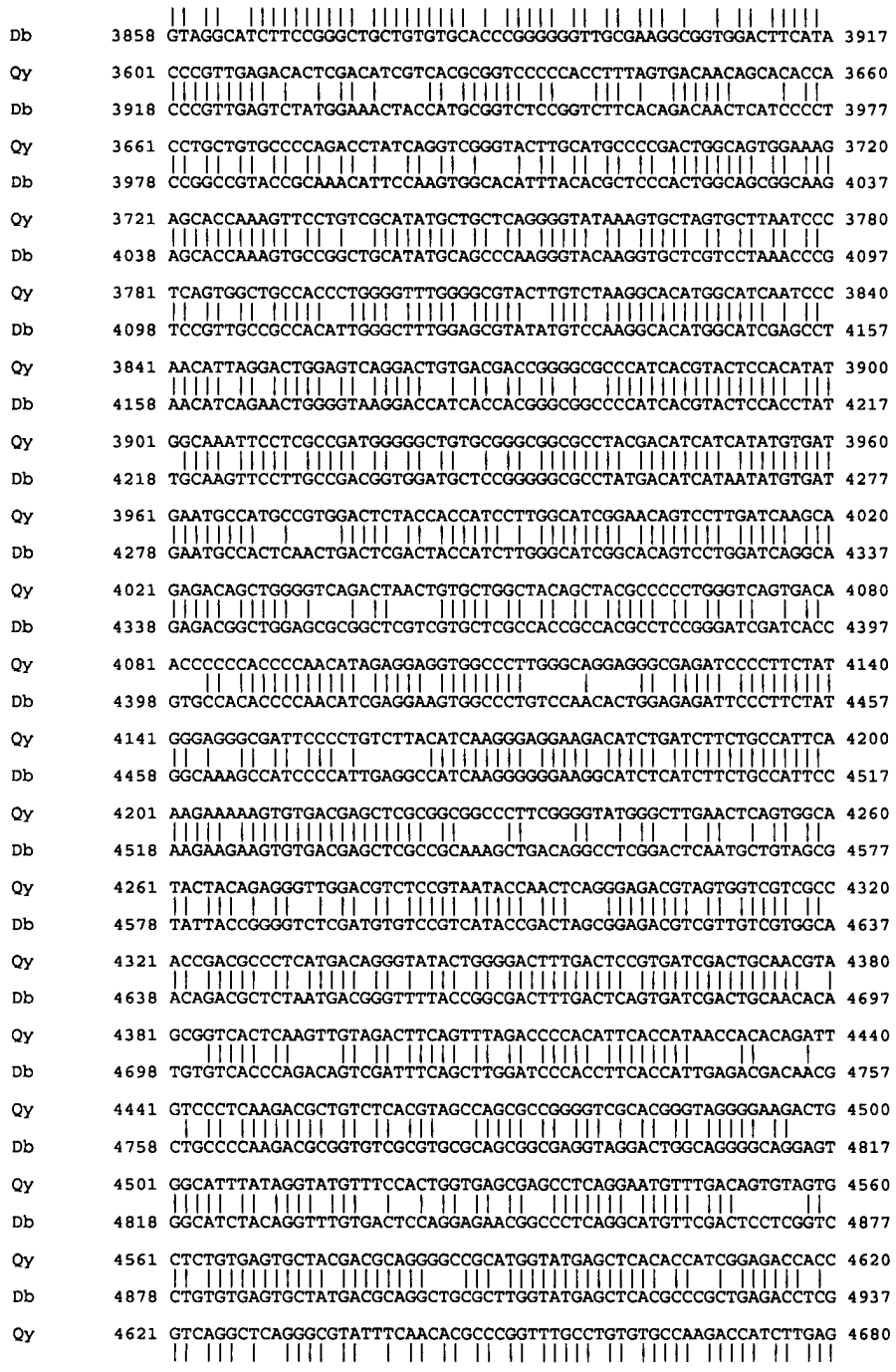
Qy       1141 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1200
              || |||| || ||||||      |||||   |  ||    | ||||  | |||||
Db       1470 GTTGACGGGCACACCCACGTGACAGGGGGAAGGGTAGCCTCCAGCACCCAGAGCCTCGTG 1529

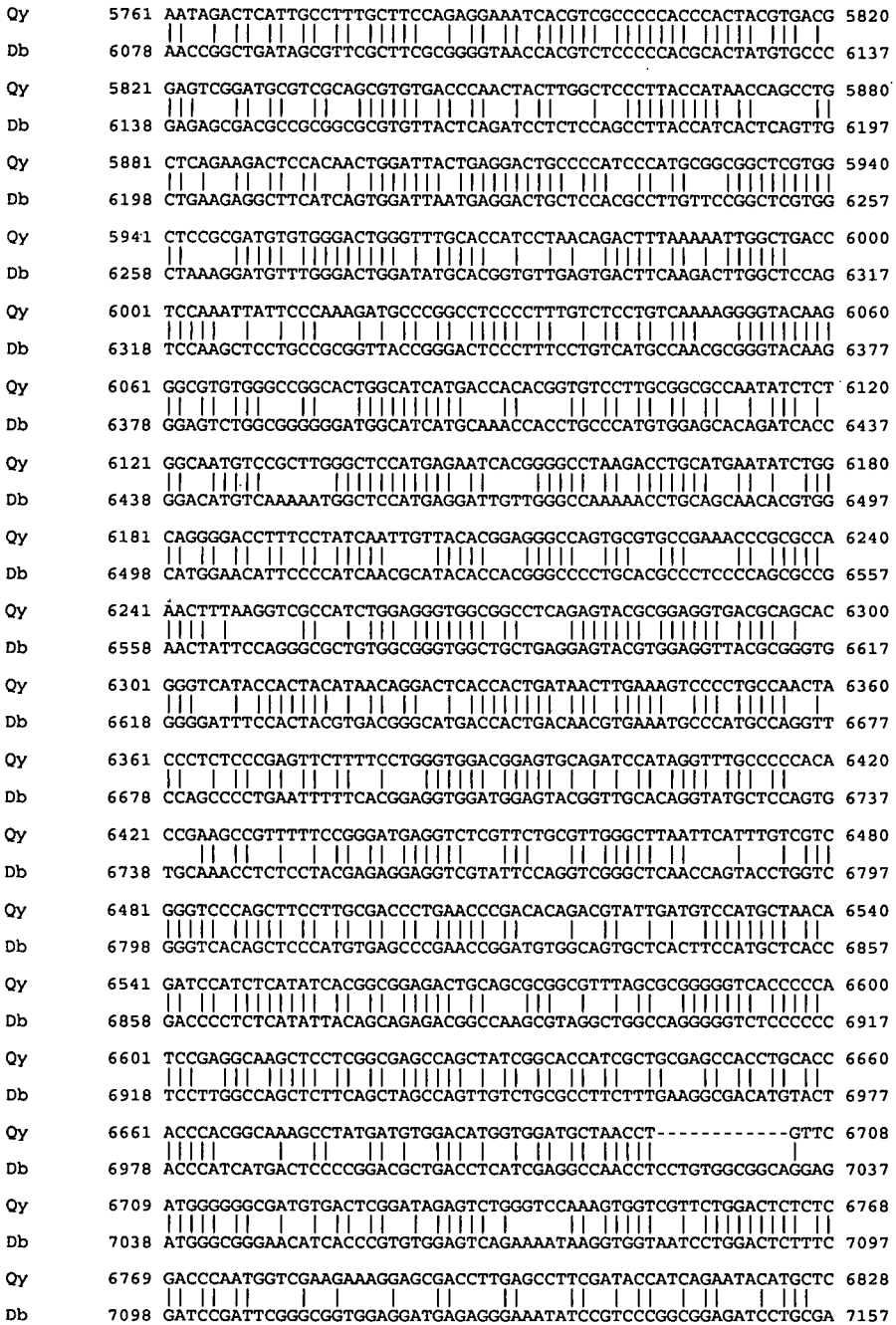
Qy       1201 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1260
              ||    |     |||||   |||||||||| ||||| ||||||| |||||||||
Db       1530 TCCTGGCTCTCACAAGGCCCATCTCAGAAAATCCAACTCGTGAACACCAACGGCAGCTGG 1589

Qy       1261 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1320
              ||||||||| | ||||| ||||| |||||||||||| | || || || || || || |
Db       1590 CACATCAACAGGACCGCTCTGAATTGCAATGACTCCCTCCAAACTGGGTTCATTGCTGCG 1649

Qy       1321 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1380
              ||||||||| | ||||| |||||| |||| || || || || |||||| |   |||||||
Db       1650 CTGTTCTACGCACACAGGTTCAACGCGTCCGGGTGCCCAGAGCGCATGGCTAGCTGCCGC 1709
```


Qy 1381 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1440
Db 1710 CCCATCGATGAGTTCGCTCAGGGGTGGGGTCCCATCACTCATGATATGCCTGAGAGCTC- 1768
Qy 1441 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1500
Db 1769 -----GGACCAGAGGCCATATTGCTGGCACTACGCGCCTCGACCGTGCGGGATCGTGCCT 1823
Qy 1501 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1560
Db 1824 GCGTCGCAGGTGTGTGGTCCAGTGTATTGCTTCACTCCGAGCCCTGTTGTAGTGGGGACG 1883
Qy 1561 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1620
Db 1884 ACCGATCGTTTCGGCGCTCCTACGTATAGCTGGGGGGAGAATGAGACAGACGTGCTGCTA 1943
Qy 1621 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 1680
Db 1944 CTTAGCAACACGCGGCCGCCTCAAGGCAACTGGTTTGGGTGCACGTGGATGAACAGCACT 2003
Qy 1681 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 1740
Db 2004 GGGTTCACCAAGACGTGCGGGGGCCCTCCGTGCAACATCGGGGGGGTCGGCAACAACA-- 2061
Qy 1741 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 1800
Db 2062 ----CCTTGGTCTGCCCCACGGATTGCTTCCGGAAGCACCCCGAGGCCACTTACACAAAG 2117
Qy 1801 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 1860
Db 2118 TGTGGCTCGGGGCCCTGGTTGACACCCAGGTGCATGGTTGACTACCCATACAGGCTCTGG 2177
Qy 1861 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 1920
Db 2178 CACTACCCCTGCACTGTTAACTTTACCGTCTTTAAGGTCAGGATGTATGTGGGGGGCGTG 2237
Qy 1921 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 1980
Db 2238 GAGCACAGGCTCAATGCTGCATGCAATTGGACTCGAGGAGAGCGCTGTGACTTGGAGGAC 2297
Qy 1981 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2040
Db 2298 AGGGATAGGTCAGAACTCAGCCCGCTGCTGCTGTCTACAACAGAGTGGCAGATACTGCCC 2357
Qy 2041 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2100
Db 2358 TGTTCCTTCACCACCCTACCGGCCCTGTCCACTGGCTTGATCCATCTTCACCGGAACATC 2417
Qy 2101 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2160
Db 2418 GTGGACGTGCAATACCTGTACGGTATAGGGTCGGCAGTTGTCTCCTTTGCAATCAAATGG 2477
Qy 2161 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2220
Db 2478 GAGTATATCCTGTTGCTTTTCCTTCTTCTGGCGGACGCGCGCGTCTGTGCCTGCTTGTGG 2537
Qy 2221 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2280
Db 2538 ATGATGCTGCTGATAGCCCAGGCTGAGGCCACCTTAGAGAACCTGGTGGTCCTCAATGCG 2597
Qy 2281 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATATTTTGTCATCTTTTTCGTGGCTGCTTGG 2340
Db 2598 GCGTCTGTGGCCGGAGCGCATGGCCTTCTCTCCTTCCTCGTGTTCTTCTGCGCCGCCTGG 2657
Qy 2341 TACATCAAGGGTCGGGTAGTCCCCTTAGCTACCTATTCCCTCACTGGCCTGTGGTCCTTT 2400
Db 2658 TACATCAAAGGCAGGCTGGTCCCTGGGGCGGCATATGCTCTCTATGGCGTATGGCCGTTG 2717
Qy 2401 AGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTGCATGGC 2460
Db 2718 CTCCTGCTCTTGCTGGCCTTACCACCACGAGCTTATGCCATGGACCGAGAGATGGCTGCA 2777
Qy 2461 CAGATAGGAGCGGCTCTGCTGGTAATGATCACTCTCTTTACTCTCACCCCCGGGTATAAG 2520
Db 2778 TCGTGCGGAGGCGCGGTTTTTGTAGGTCTGGTACTCTTGACCTTGTCACCATACTATAAG 2837
Qy 2521 ACCCTTCTCAGCCGGTTTTTGTGGTGGTTGTGCTATCTTCTGACCCTGGGGGAAGCTATG 2580
Db 2838 GTGTTCCTCGCTAGGCTCATATGGTGGTTACAATATTTTATCACCAGAGCCGAGGCGCAC 2897
Qy 2581 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGCATCATATGGGCC 2640
Db 2898 TTGCAAGTGTGGGTCCCCCCTCTCAATGTTCGGGGAGGCCGCGATGCCATCATCCTCCTT 2957
Qy 2641 GTCGCCATATTCTACCCAGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 2700
Db 2958 ACATGCGCGGTCCATCCAGAGCTAATCTTTGACATCACCAAACTCCTGCTCGCCATACTC 3017
Qy 2701 GGGCCTGCTTACCTCCTAAAAGGTGCTTTGACGCGCGTGCCGTACTTCGTCAGGGCTCAC 2760
Db 3018 GGTCCGCTCATGGTGCTCCAGGCTGGCATAACTAGAGTGCCGTACTTTGTACGCGCTCAG 3077
Qy 2761 GCTCTACTGAGGATGTGCACCATGGCAAGGCATCTCGCGGGGGGCAGGTACGTCCAGATG 2820
Db 3078 GGGCTCATCCGTGCATGCATGTTAGTGCGGAAGGTCGCTGGAGGCCACTATGTCCAAATG 3137
Qy 2821 GCGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 2880
Db 3138 GCCTTCATGAAGCTGGCCGCGCTGACAGGTACGTACGTATATGACCATCTTACTCCACTG 3197
Qy 2881 TCGGATTGGGCTGCTAGTGGCCTGCGGGACCTGGCGGTCGCCGTTGAGCCTATCATCTTC 2940
Db 3198 CGGGATTGGGCCCACGCGGGCCTACGAGACCTTGCGGTGGCAGTAGAGCCCGTCGTCTTC 3257
Qy 2941 AGTCCGATGGAGAAGAAAGTCATTGTCTGGGGAGCGGAGACAGCTGCTTGTGGGGACATT 3000
Db 3258 TCTGACATGGAGACTAAACTCATCACCTGGGGGGCAGACACCGCGGCGTGTGGGGACATC 3317
Qy 3001 TTACACGGACTTCCCGTGTCCGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3060
Db 3318 ATCTCGGGTCTACCAGTCTCCGCCCGAAGGGGGAAGGAGATACTTCTAGGACCGGCCGAT 3377
Qy 3061 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3120
Db 3378 AGTTTTGGAGAGCAGGGGTGGCGGCTCCTTGCGCCTATCACGGCCTATTCCCAACAAACG 3437
Qy 3121 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCT 3180
Db 3438 CGGGGCCTGCTTGGCTGTATCATCACTAGCCTCACAGGTCGGGACAAGAACCAGGTCGAT 3497
Qy 3181 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3240
Db 3498 GGGGAGGTTCAGGTGCTCTCCACCGCAACGCAATCTTTCCTGGCGACCTGCGTCAATGGC 3557
Qy 3241 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3300
Db 3558 GTGTGTTGGACCGTCTACCATGGTGCCGGCTCGAAGACCCTGGCCGGCCCGAAGGGTCCA 3617
Qy 3301 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGG 3360
Db 3618 ATCACCCAAATGTACACCAATGTAGACCAGGACCTCGTCGGCTGGCCGGCGCCCCCCGGG 3677
Qy 3361 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3420
Db 3678 GCGCGCTCCATGACACCGTGCACCTGCGGCAGCTCGGACCTTTACTTGGTCACGAGGCAT 3737
Qy 3421 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3480
Db 3738 GCTGATGTCGTTCCGGTGCGCCGGCGGGGCGACAGCAGGGGGAGCCTGCTTTCCCCCAGG 3797
Qy 3481 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3540
Db 3798 CCCATCTCCTACCTGAAGGGCTCCTCGGGTGGACCACTGCTTTGCCCTTCGGGGCACGTT 3857
Qy 3541 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3600


Db 3858 GTAGGCATCTTCCGGGCTGCTGTGTGCACCCGGGGGGTTGCGAAGGCGGTGGACTTCATA 3917
Qy 3601 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 3660
Db 3918 CCCGTTGAGTCTATGGAAACTACCATGCGGTCTCCGGTCTTCACAGACAACTCATCCCCT 3977
Qy 3661 CCTGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 3720
Db 3978 CCGGCCGTACCGCAAACATTCCAAGTGGCACATTTACACGCTCCCACTGGCAGCGGCAAG 4037
Qy 3721 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 3780
Db 4038 AGCACCAAAGTGCCGGCTGCATATGCAGCCCAAGGGTACAAGGTGCTCGTCCTAAACCCG 4097
Qy 3781 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 3840
Db 4098 TCCGTTGCCGCCACATTGGGCTTTGGAGCGTATATGTCCAAGGCACATGGCATCGAGCCT 4157
Qy 3841 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 3900
Db 4158 AACATCAGAACTGGGGTAAGGACCATCACCACGGGCGGCCCCATCACGTACTCCACCTAT 4217
Qy 3901 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 3960
Db 4218 TGCAAGTTCCTTGCCGACGGTGGATGCTCCGGGGGCGCCTATGACATCATAATATGTGAT 4277
Qy 3961 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4020
Db 4278 GAATGCCACTCAACTGACTCGACTACCATCTTGGGCATCGGCACAGTCCTGGATCAGGCA 4337
Qy 4021 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4080
Db 4338 GAGACGGCTGGAGCGCGGCTCGTCGTGCTCGCCACCGCCACGCCTCCGGGATCGATCACC 4397
Qy 4081 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4140
Db 4398 GTGCCACACCCCAACATCGAGGAAGTGGCCCTGTCCAACACTGGAGAGATTCCCTTCTAT 4457
Qy 4141 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4200
Db 4458 GGCAAAGCCATCCCCATTGAGGCCATCAAGGGGGGAAGGCATCTCATCTTCTGCCATTCC 4517
Qy 4201 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4260
Db 4518 AAGAAGAAGTGTGACGAGCTCGCCGCAAAGCTGACAGGCCTCGGACTCAATGCTGTAGCG 4577
Qy 4261 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4320
Db 4578 TATTACCGGGGTCTCGATGTGTCCGTCATACCGACTAGCGGAGACGTCGTTGTCGTGGCA 4637
Qy 4321 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4380
Db 4638 ACAGACGCTCTAATGACGGGTTTTACCGGCGACTTTGACTCAGTGATCGACTGCAACACA 4697
Qy 4381 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4440
Db 4698 TGTGTCACCCAGACAGTCGATTTCAGCTTGGATCCCACCTTCACCATTGAGACGACAACG 4757
Qy 4441 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4500
Db 4758 CTGCCCCAAGACGCGGTGTCGCGTGCGCAGCGGCGAGGTAGGACTGGCAGGGGCAGGAGT 4817
Qy 4501 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4560
Db 4818 GGCATCTACAGGTTTGTGACTCCAGGAGAACGGCCCTCAGGCATGTTCGACTCCTCGGTC 4877
Qy 4561 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4620
Db 4878 CTGTGTGAGTGCTATGACGCAGGCTGCGCTTGGTATGAGCTCACGCCCGCTGAGACCTCG 4937
Qy 4621 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 4680
Db 4938 GTTAGGTTGCGGGCTTACCTAAATACACCAGGGTTGCCCGTCTGCCAGGACCACCTAGAG 4997
Qy 4681 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 4740
Db 4998 TTCTGGGAGAGCGTCTTCACAGGCCTCACCCACATAGATGCCCACTTCTTGTCCCAGACC 5057
Qy 4741 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 4800
Db 5058 AAACAGGCAGGAGACAACCTCCCCTACCTGGTAGCATACCAAGCCACAGTGTGCGCCAGG 5117
Qy 4801 GCCAAAGCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 4860
Db 5118 GCTCAGGCTCCACCTCCATCGTGGGACCAAATGTGGAAGTGTCTCATACGGCTAAAGCCC 5177
Qy 4861 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 4920
Db 5178 ACACTGCATGGGCCAACGCCCCTGCTGTACAGGCTAGGAGCCGTTCAAAATGAGGTCACT 5237
Qy 4921 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 4980
Db 5238 CTCACACACCCCATAACCAAATACATCATGGCATGCATGTCGGCTGACCTGGAGGTCGTC 5297
Qy 4981 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5040
Db 5298 ACTAGCACCTGGGTGCTAGTAGGCGGAGTCCTTGCGGCTCTGGCCGCGTACTGCCTGACG 5357
Qy 5041 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5100
Db 5358 ACAGGCAGCGTGGTCATTGTGGGCAGGATCATCTTGTCCGGGAGGCCAGCTGTTATTCCC 5417
Qy 5101 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5160
Db 5418 GACAGGGAAGTCCTCTACCAGGAGTTCGATGAGATGGAAGAGTGTGCTTCACACCTCCCT 5477
Qy 5161 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5220
Db 5478 TACATCGAGCAAGGAATGCAGCTCGCCGAGCAATTCAAACAGAAGGCGCTCGGATTGCTG 5537
Qy 5221 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5280
Db 5538 CAAACAGCCACCAAGCAAGCGGAGGCTGCTGCTCCCGTGGTGGAGTCCAAGTGGCGAGCC 5597
Qy 5281 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5340
Db 5598 CTTGAGGTCTTCTGGGCGAAACACATGTGGAACTTCATCAGCGGGATACAGTACTTGGCA 5657
Qy 5341 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5400
Db 5658 GGCCTATCCACTCTGCCTGGAAACCCCGCGATAGCATCATTGATGGCTTTTACAGCCTCT 5717
Qy 5401 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5460
Db 5718 ATCACCAGCCCGCTCACCACCCAAAATACCCTCCTGTTTAACATCTTGGGGGGATGGGTG 5777
Qy 5461 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5520
Db 5778 GCTGCCCAACTCGCTCCCCCCAGCGCTGCTTCGGCTTTCGTGGGCGCCGGCATTGCCGGT 5837
Qy 5521 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5580
Db 5838 GCGGCCGTTGGCAGCATAGGTCTCGGGAAGGTACTTGTGGACATTCTGGCGGGCTATGGG 5897
Qy 5581 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5640
Db 5898 GCGGGGGTGGCTGGCGCACTCGTGGCCTTTAAGGTCATGAGCGGCGAGATGCCCTCCACT 5957
Qy 5641 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 5700
Db 5958 GAGGATCTGGTTAATTTACTCCCTGCCATCCTTTCTCCTGGCGCCCTGGTTGTCGGGGTC 6017
Qy 5701 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 5760
Db 6018 GTGTGCGCAGCAATACTGCGTCGGCACGTGGGCCCGGGAGAGGGGGCTGTGCAGTGGATG 6077


Qy 5761 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 5820
Db 6078 AACCGGCTGATAGCGTTCGCTTCGCGGGGTAACCACGTCTCCCCCACGCACTATGTGCCC 6137
Qy 5821 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 5880
Db 6138 GAGAGCGACGCCGCGGCGCGTGTTACTCAGATCCTCTCCAGCCTTACCATCACTCAGTTG 6197
Qy 5881 CTCAGAAGACTCCACAACTGGATTACTGAGGACTGCCCCATCCCATGCGGCGGCTCGTGG 5940
Db 6198 CTGAAGAGGCTTCATCAGTGGATTAATGAGGACTGCTCCACGCCTTGTTCCGGCTCGTGG 6257
Qy 5941 CTCCGCGATGTGTGGGACTGGGTTTGCACCATCCTAACAGACTTTAAAAATTGGCTGACC 6000
Db 6258 CTAAAGGATGTTTGGGACTGGATATGCACGGTGTTGAGTGACTTCAAGACTTGGCTCCAG 6317
Qy 6001 TCCAAATTATTCCCAAAGATGCCCGGCCTCCCCTTTGTCTCCTGTCAAAAGGGGTACAAG 6060
Db 6318 TCCAAGCTCCTGCCGCGGTTACCGGGACTCCCTTTCCTGTCATGCCAACGCGGGTACAAG 6377
Qy 6061 GGCGTGTGGGCCGGCACTGGCATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6120
Db 6378 GGAGTCTGGCGGGGGGATGGCATCATGCAAACCACCTGCCCATGTGGAGCACAGATCACC 6437
Qy 6121 GGCAATGTCCGCTTGGGCTCCATGAGAATCACGGGGCCTAAGACCTGCATGAATATCTGG 6180
Db 6438 GGACATGTCAAAAATGGCTCCATGAGGATTGTTGGGCCAAAAACCTGCAGCAACACGTGG 6497
Qy 6181 CAGGGGACCTTTCCTATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCGCCA 6240
Db 6498 CATGGAACATTCCCCATCAACGCATACACCACGGGCCCCTGCACGCCCTCCCCAGCGCCG 6557
Qy 6241 AACTTTAAGGTCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6300
Db 6558 AACTATTCCAGGGCGCTGTGGCGGGTGGCTGCTGAGGAGTACGTGGAGGTTACGCGGGTG 6617
Qy 6301 GGGTCATACCACTACATAACAGGACTCACCACTGATAACTTGAAAGTCCCCTGCCAACTA 6360
Db 6618 GGGGATTTCCACTACGTGACGGGCATGACCACTGACAACGTGAAATGCCCATGCCAGGTT 6677
Qy 6361 CCCTCTCCCGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCACA 6420
Db 6678 CCAGCCCCTGAATTTTTCACGGAGGTGGATGGAGTACGGTTGCACAGGTATGCTCCAGTG 6737
Qy 6421 CCGAAGCCGTTTTTCCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6480
Db 6738 TGCAAACCTCTCCTACGAGAGGAGGTCGTATTCCAGGTCGGGCTCAACCAGTACCTGGTC 6797
Qy 6481 GGGTCCCAGCTTCCTTGCGACCCTGAACCCGACACAGACGTATTGATGTCCATGCTAACA 6540
Db 6798 GGGTCACAGCTCCCATGTGAGCCCGAACCGGATGTGGCAGTGCTCACTTCCATGCTCACC 6857
Qy 6541 GATCCATCTCATATCACGGCGGAGACTGCAGCGCGGCGTTTAGCGCGGGGGTCACCCCCA 6600
Db 6858 GACCCCTCTCATATTACAGCAGAGACGGCCAAGCGTAGGCTGGCCAGGGGGTCTCCCCCC 6917
Qy 6601 TCCGAGGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 6660
Db 6918 TCCTTGGCCAGCTCTTCAGCTAGCCAGTTGTCTGCGCCTTCTTTGAAGGCGACATGTACT 6977
Qy 6661 ACCCACGGCAAAGCCTATGATGTGGACATGGTGGATGCTAACCT----------GTTC 6708
Db 6978 ACCCATCATGACTCCCCGGACGCTGACCTCATCGAGGCCAACCTCCTGTGGCGGCAGGAG 7037
Qy 6709 ATGGGGGGCGATGTGACTCGGATAGAGTCTGGGTCCAAAGTGGTCGTTCTGGACTCTCTC 6768
Db 7038 ATGGGCGGGAACATCACCCGTGTGGAGTCAGAAAATAAGGTGGTAATCCTGGACTCTTTC 7097
Qy 6769 GACCCAATGGTCGAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCAGAATACATGCTC 6828
Db 7098 GATCCGATTCGGGCGGTGGAGGATGAGAGGGAAATATCCGTCCCGGCGGAGATCCTGCGA 7157
Qy 6829 CCCAAGAAGAGGTTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCG 6888
Db 7158 AAACCCAGGAAGTTCCCCCCAGCGTTGCCCATATGGGCACGCCCGGATTACAACCCTCCA 7217
Qy 6889 CTTGTGGAATCGTGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTC 6948
Db 7218 CTGCTAGAGTCCTGGAAGGACCCGGACTACGTCCCCCCGGTGGTACACGGGTGCCCTTTG 7277
Qy 6949 CCTCCTCCTAGGAAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGT 7008
Db 7278 CCATCTACCAAGGCCCCCCCAATACCACCTCCACGGAGGAAGAGGACGGTTGTCCTGACA 7337
Qy 7009 GAGGACTCCATAGGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCC 7068
Db 7338 GAGTCCACCGTGTCTTCTGCCTTGGCGGAGCTCGCTACTAAGACCTTTGGCAGCTCCGGG 7397
Qy 7069 CCAAGCGGCGATTCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACG 7128
Db 7398 TCGTCGGCCGTTGACAGCGG---CACGGCGACTGGCCCTCCCGA---------TCAGGCC 7445
Qy 7129 CCTCCTGATGAGTTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCTCGAG 7188
Db 7446 TCCGACGACGGCGACAAAGGATCCGACGTTGAGTCGTACTCCTCCATGCCCCCCCTCGAG 7505
Qy 7189 GGGGAGCTTGGAGATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCAG 7248
Db 7506 GGAGAGCCAGGGGACCCCGACCT---------- 7528
Qy 7249 GGGGGGGTGGCAGCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGAC 7308
Db 7529 ----------CAGCGACGGGTCTTGGTCTACCGTGAGCGGGGAAGCTGGT 7568
Qy 7309 GACTCCGTCGTGTGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGT 7368
Db 7569 GAGGACGTCGTCTGCTGCTCAATGTCCTATACATGGACAGGTGCCTTGATCACGCCATGC 7628
Qy 7369 AGTCCCGAAGAGGAGAAGTTACCGATTAACCCCTTGAGCAACTCCCTGTTGCGATATCAC 7428
Db 7629 GCTGCGGAGGAGAGCAAGTTGCCCATCAATCCGTTGAGCAACTCTTTGCTGCGTCACCAC 7688
Qy 7429 AACAAGGTGTACTGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTT 7488
Db 7689 AGTATGGTCTACTCCACAACATCTCGCAGCGCAAGTCTGCGGCAGAAGAAGGTCACCTTT 7748
Qy 7489 GATAGGATGCAAGTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCG 7548
Db 7749 GACAGACTGCAAGTCCTGGACGACCACTACCGGGACGTGCTCAAGGAGATGAAGGCGAAG 7808
Qy 7549 GCCTCCAAGGTCACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCC 7608
Db 7809 GCGTCCACAGTTAAGGCTAGGCTTCTATCTATAGAGGAGGCCTGCAAACTGACGCCCCCA 7868
Qy 7609 CATTCTGCAAGATCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGG 7668
Db 7869 CATTCGGCCAAATCCAAATTTGGCTACGGGGCGAAGGACGTCCGGAGCCTATCCAGCAGG 7928
Qy 7669 GCCGTTAACCACATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATT 7728
Db 7929 GCCGTCAACCACATCCGCTCCGTGTGGGAGGACTTGCTGGAAGACACTGAAACACCAATT 7988
Qy 7729 CCCACAACCATTATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAG 7788
Db 7989 GATACCACCATCATGGCAAAAAATGAGGTTTTCTGCGTCCAACCAGAGAAAGGAGGCCGC 8048
Qy 7789 AAAGCAGCTCGCCTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCC 7848
Db 8049 AAGCCAGCTCGCCTTATCGTATTCCCAGACCTGGGGGTACGTGTATGCGAGAAGATGGCC 8108
Qy 7849 CTTTATGACATTACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAG 7908
Db 8109 CTTTACGACGTGGTCTCCACCCTTCCTCAGGCCGTGATGGGCCCCTCATACGGATTCCAG 8168
Qy 7909 TATTCCCCCGCTCAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCT 7968

```
Db       8169 TACTCTCCTGGGCAGCGGGTCGAGTTCCTGGTGAATACCTGGAAATCAAAGAAATGCCCT 8228

Qy       7969 ATGGGTTTTTCGTATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGG 8028

Db       8229 ATGGGCTTCTCATATGACACCCGCTGCTTTGACTCAACGGTCACTGAGAATGACATCCGT 8288

Qy       8029 ACTGAGGAGTCCATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACAC 8088

Db       8289 ACTGAGGAATCAATTTACCAATGTTGTGACTTGGCCCCCGAAGCCAGGCAGGCCATAAGG 8348

Qy       8089 TCGCTAACTGAGAGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGC 8148

Db       8349 TCGCTCACAGAGCGGCTTTATGTCGGGGGTCCCCTGACTAATTCGAAGGGGCAGAACTGC 8408

Qy       8149 GGGTACAGGCGTTGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACA 8208

Db       8409 GGTTATCGCCGGTGCCGCGCAAGTGGCGTGCTGACGACTAGCTGCGGCAACACCCTCACA 8468

Qy       8209 TGCTACGTGAAAGCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTG 8268

Db       8469 TGTTACTTGAAGGCCACTGCGGCCTGTCGAGCTGCAAAGCTCCAGGACTGCACGATGCTC 8528

Qy       8269 GTATGCGGCGATGACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGG 8328

Db       8529 GTGAACGGAGACGACCTTGTCGTTATCTGTGAGAGTGCGGGAACCCAGGAGGATGCGGCG 8588

Qy       8329 AACCTGAGAGCCTTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCC 8388

Db       8589 GCCCTACGAGCCTTCACGGAGGCTATGACTAGGTATTCCGCCCCCCCCGGGGACCCGCCC 8648

Qy       8389 AGACCGGAGTATGATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTG 8448

Db       8649 CAACCAGAATACGACTTGGAGCTGATAACGTCATGCTCCTCCAATGTGTCGGTCGCGCAC 8708

Qy       8449 GGCCCACAAGGCCGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGG 8508

Db       8709 GATGCATCCGGCAAAAGGGTGTACTACCTCACCCGTGACCCCACCACCCCCCTCGCACGG 8768

Qy       8509 GCTGCCTGGGAAACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAG 8568

Db       8769 GCTGCGTGGGAGACAGTTAGACACACTCCAGTCAACTCCTGGCTAGGCAATATCATCATG 8828

Qy       8569 TACGCCCCGACCATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATG 8628

Db       8829 TATGCGCCCACCCTATGGGCGAGGATGATTCTGATGACTCATTTCTTCTCTATCCTTCTA 8888

Qy       8629 GCTCAAGACACGCTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTG 8688

Db       8889 GCTCAGGAGCAACTTGAAAAAGCCCTGGATTGTCAGATCTACGGGGCCTGTTACTCCATT 8948

Qy       8689 AGTCCCTTGGACCTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTG 8748

Db       8949 GAGCCACTTGACCTACCTCAGATCATTGAACGACTCCATGGTCTTAGCGCATTTTCACTC 9008

Qy       8749 CACACATACACTCCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCG 8808

Db       9009 CACAGTTACTCTCCAGGTGAGATCAATAGGGTGGCTTCATGCCTCAGGAAACTTGGGGTA 9068

Qy       8809 CCACCCCTCAGAGCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGT 8868

Db       9069 CCGCCTTTGCGAGTCTGGAGACATCGGGCCAGAAGTGTCCGCGCTAAGCTACTGTCCCAG 9128

Qy       8869 GGGGGGAGAGCGGCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTC 8928

Db       9129 GGGGGGAGGGCTGCCACTTGCGGCAAGTACCTCTTCAACTGGGCAGTAAAGACCAAGCTT 9188

Qy       8929 AAACTCACTCCATTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGC 8988

Db       9189 AAACTCACTCCAATCCCGGCTGCGTCCCAGCTAGACTTGTCCGGCTGGTTCGTTGCTGGT 9248

Qy       8989 GCCGGCGGGGGCGACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTT 9048

Db       9249 TACAACGGGGGAGACATATATCACAGCCTGTCTCGTGCCCGACCCCGTTGGTTCATGTTG 9308

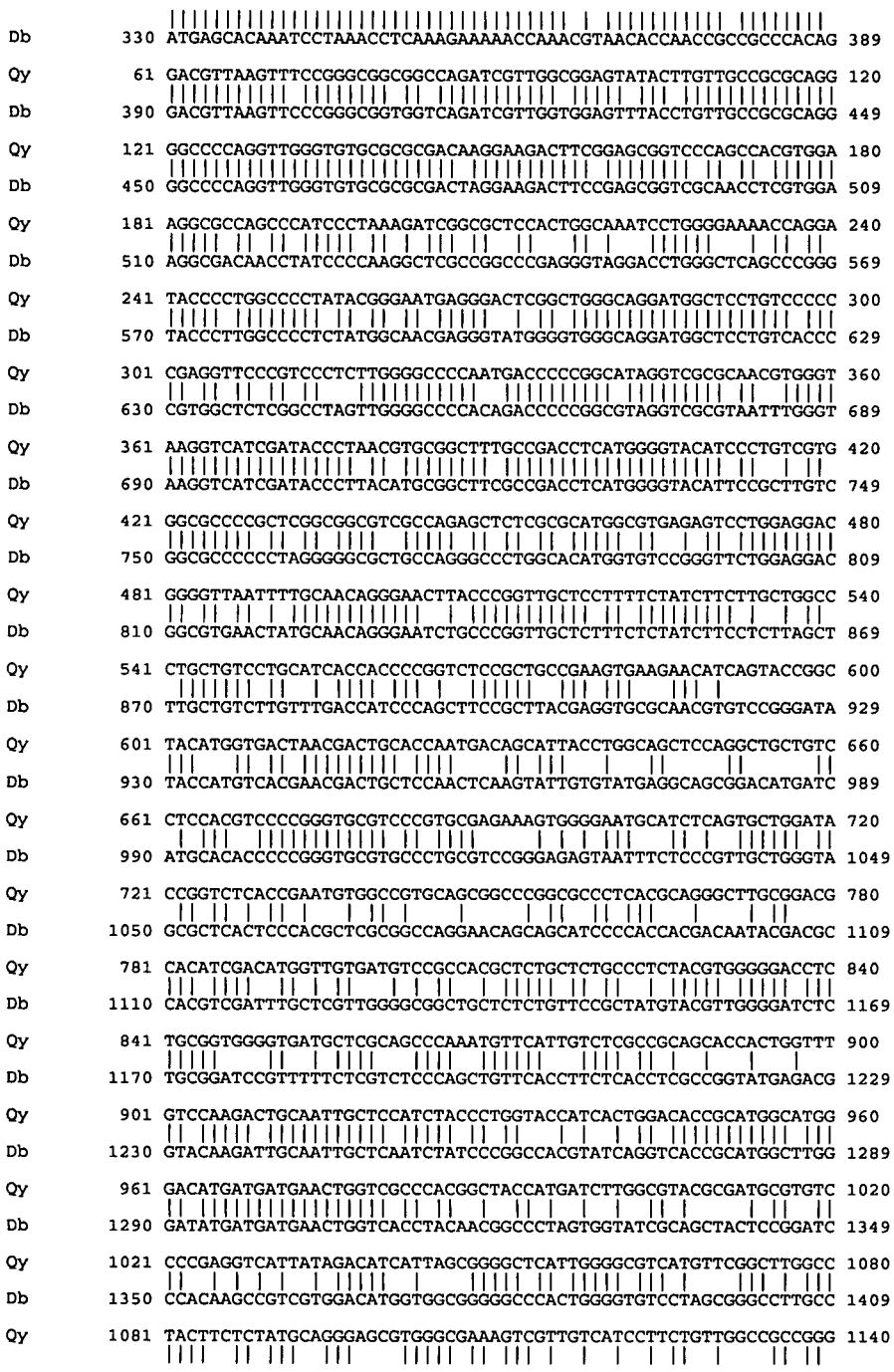
Qy       9049 GGCCTACTCCTACTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCGG 9099

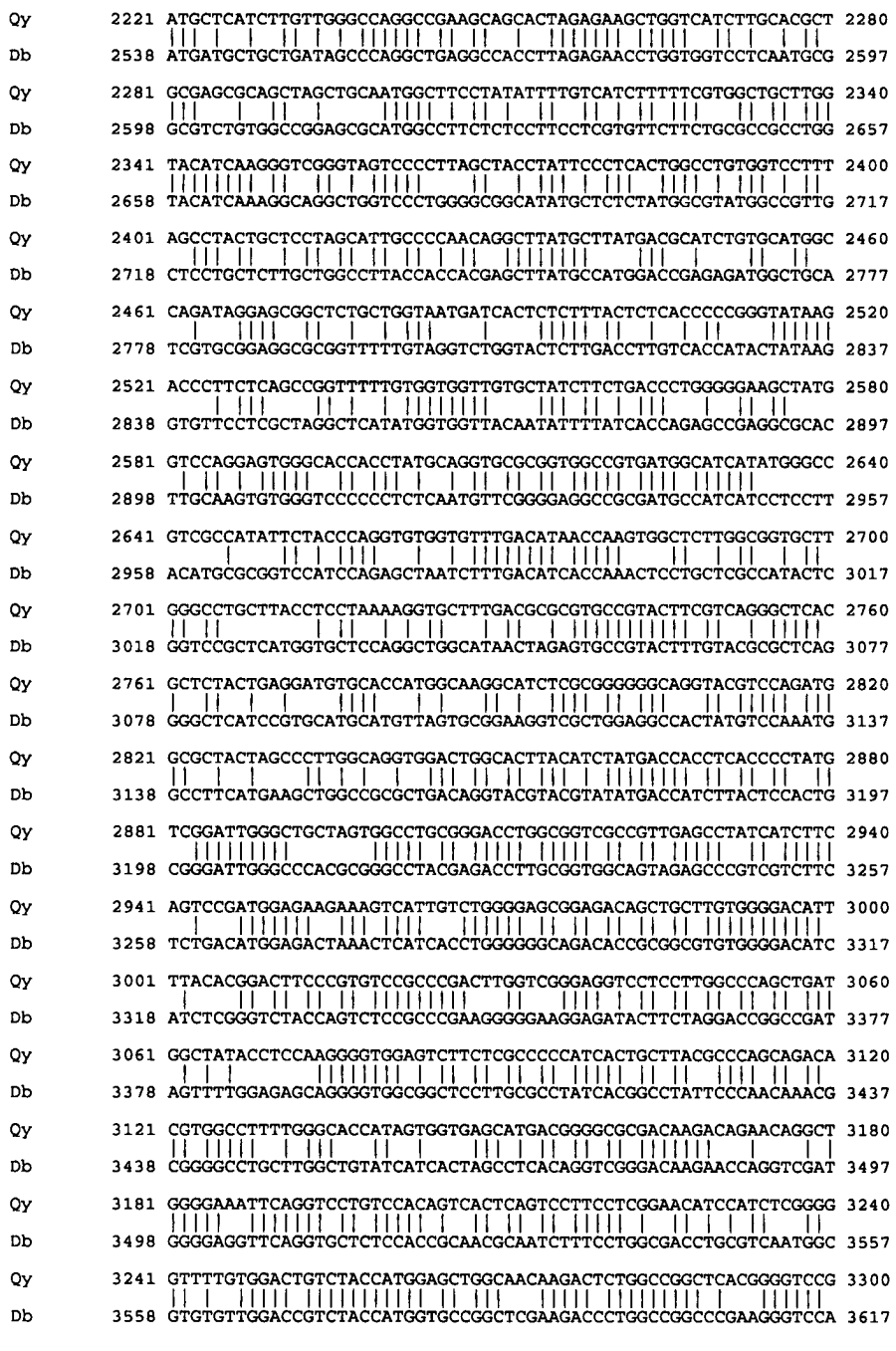
Db       9309 TGCCTACTCCTACTTTCTGTAGGGGTAGGCATCTACCTGCTCCCCAACCGG 9359
```

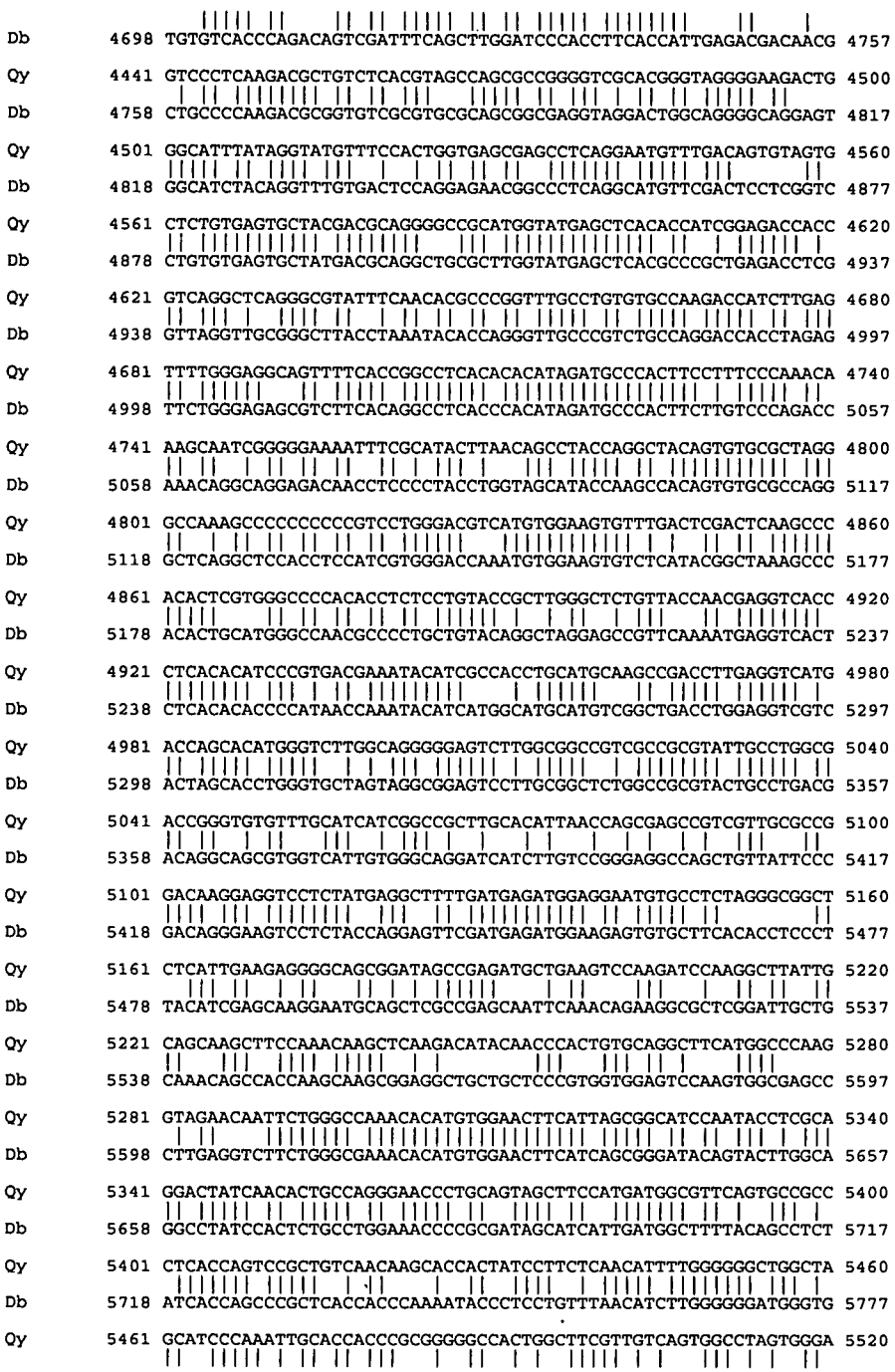
RESULT 13
AAD49655
ID AAD49655 standard; DNA; 9413 BP.
XX
AC AAD49655;
XX
DT 24-MAR-2003 (first entry)
XX
DE Hepatitis C virus (HCV) DNA.
XX
KW Amyloidosis; haemophilia; Alzheimer's disease; atherosclerosis; cancer;
KW gigantism; dwarfism; hypothyroidism; hyperthyroidism; cystic fibrosis;
KW autoimmune disorder; aging; inflammation; diabetes; obesity; anorectic;
KW neurodegenerative disorder; Parkinson's disease; gene therapy; virucide;
KW haemostatic; antibacterial; nootropic; neuroprotective; cytostatic;
KW fungicide; Hepatitis C virus; HCV; gene; ds.
XX
OS Hepatitis C virus.
XX
PN WO200283953-A1.
XX
PD 24-OCT-2002.
XX
PF 11-APR-2002; 2002WO-US011757.
XX
PR 11-APR-2001; 2001US-0282965P.
XX
PA (PTCT-) PTC THERAPEUTICS INC.
XX
PI Rando R, Welch E;
XX
DR WPI; 2003-075561/07.
XX
PT Identifying a test compound that binds to a target RNA molecule for
PT treating or preventing amyloidosis, hemophilia, cancer, gigantism,
PT diabetes, by contacting a detectably labeled target RNA molecule with a
PT library of test compounds.
XX
PS Example; Page 64-68; 152pp; English.
XX
CC The invention relates to a method for identifying a test compound that
CC binds to a target RNA molecule, which comprises contacting a detectably
CC labelled target RNA molecule with a library of test compounds under
CC conditions that permit direct binding of the labelled target RNA to a
CC member of the library of test compounds so that a detectably labeled
CC target RNA:test compound complex is formed. The method is useful for
CC screening libraries of compounds for those that are selectively bind to a
CC pre-selected target RNA. The compounds are useful for inhibiting the
CC formation of a specific bound RNA:host cell factor complexes in vivo.
CC They are also useful for treating or preventing diseases associated with
CC overproduction or decreased protein function, such as amyloidosis,
CC haemophilia, Alzheimer's disease, atherosclerosis, cancer, gigantism,
CC dwarfism, hypothyroidism, hyperthyroidism, autoimmune disorders, aging,
CC inflammation, cystic fibrosis, diabetes, obesity, neurodegenerative
CC disorders, Parkinson's disease or infections (bacterial, viral, fungal).
CC The invention is also used in gene therapy. The present sequence is
CC Hepatitis C virus (HCV) DNA. This sequence is used to illustrate the
CC method of the invention
XX
SQ Sequence 9413 BP; 1886 A; 2816 C; 2694 G; 2017 T; 0 U; 0 Other;

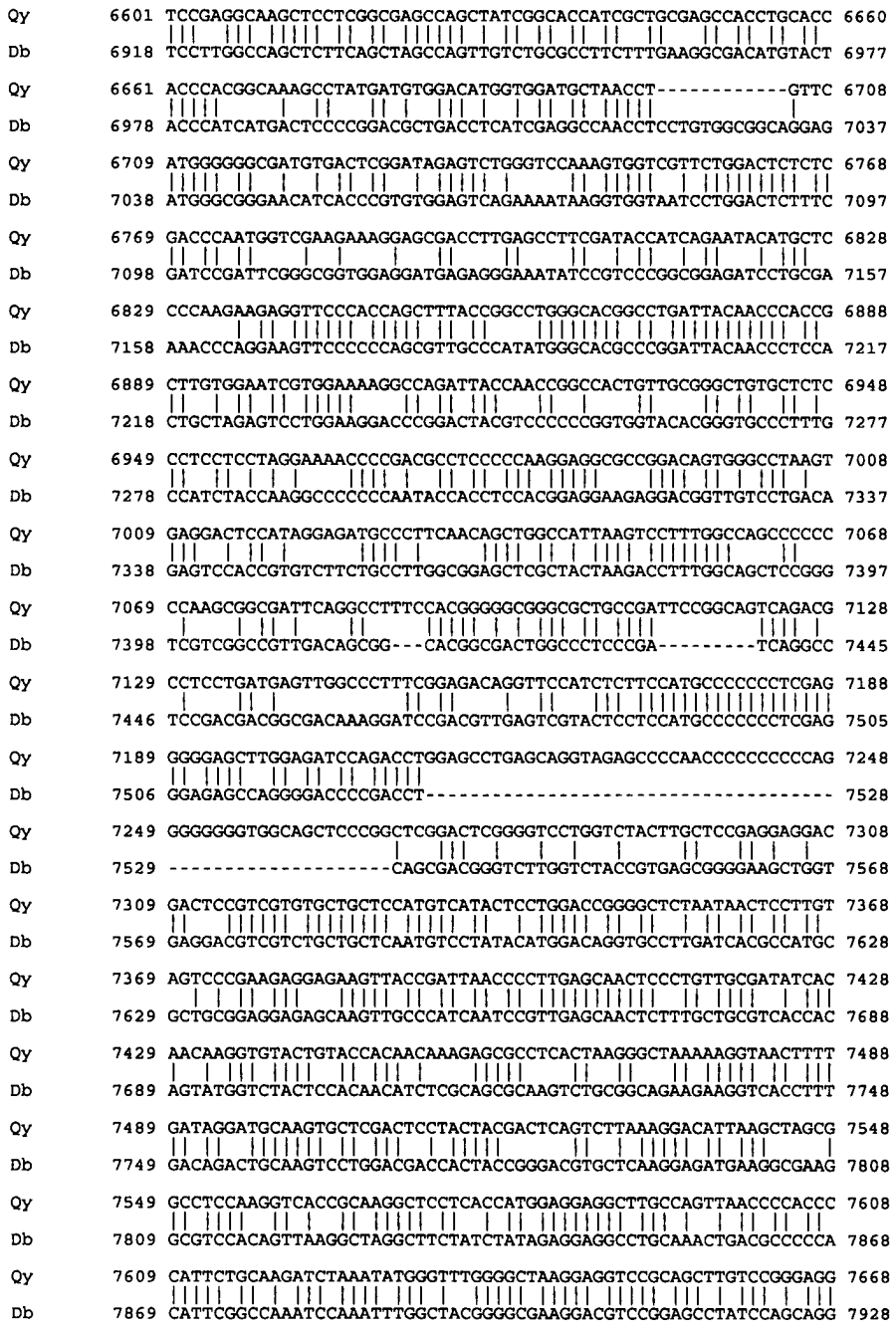
Query Match 45.5%; Score 4141.8; DB 8; Length 9413;
Best Local Similarity 66.6%; Pred. No. 0;
Matches 6066; Conservative 0; Mismatches 2952; Indels 93; Gaps 6;

```
Qy          1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 60
```


Db 330 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCCGCCCACAG 389
Qy 61 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 120
Db 390 GACGTTAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGG 449
Qy 121 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 180
Db 450 GGCCCCAGGTTGGGTGTGCGCGCGACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTGGA 509
Qy 181 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 240
Db 510 AGGCGACAACCTATCCCCAAGGCTCGCCGGCCCGAGGGTAGGACCTGGGCTCAGCCCGGG 569
Qy 241 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 300
Db 570 TACCCTTGGCCCCTCTATGGCAACGAGGGTATGGGGTGGGCAGGATGGCTCCTGTCACCC 629
Qy 301 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 360
Db 630 CGTGGCTCTCGGCCTAGTTGGGGCCCCACAGACCCCCGGCGTAGGTCGCGTAATTTGGGT 689
Qy 361 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 420
Db 690 AAGGTCATCGATACCCTTACATGCGGCTTCGCCGACCTCATGGGGTACATTCCGCTTGTC 749
Qy 421 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 480
Db 750 GGCGCCCCCCTAGGGGGCGCTGCCAGGGCCCTGGCACATGGTGTCCGGGTTCTGGAGGAC 809
Qy 481 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 540
Db 810 GGCGTGAACTATGCAACAGGGAATCTGCCCGGTTGCTCTTTCTCTATCTTCCTCTTAGCT 869
Qy 541 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 600
Db 870 TTGCTGTCTTGTTTGACCATCCCAGCTTCCGCTTACGAGGTGCGCAACGTGTCCGGGATA 929
Qy 601 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 660
Db 930 TACCATGTCACGAACGACTGCTCCAACTCAAGTATTGTGTATGAGGCAGCGGACATGATC 989
Qy 661 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 720
Db 990 ATGCACACCCCCGGGTGCGTGCCCTGCGTCCGGGAGAGTAATTTCTCCCGTTGCTGGGTA 1049
Qy 721 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 780
Db 1050 GCGCTCACTCCCACGCTCGCGGCCAGGAACAGCAGCATCCCCACCACGACAATACGACGC 1109
Qy 781 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 840
Db 1110 CACGTCGATTTGCTCGTTGGGGCGGCTGCTCTCTGTTCCGCTATGTACGTTGGGGATCTC 1169
Qy 841 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 900
Db 1170 TGCGGATCCGTTTTTCTCGTCTCCCAGCTGTTCACCTTCTCACCTCGCCGGTATGAGACG 1229
Qy 901 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 960
Db 1230 GTACAAGATTGCAATTGCTCAATCTATCCCGGCCACGTATCAGGTCACCGCATGGCTTGG 1289
Qy 961 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1020
Db 1290 GATATGATGATGAACTGGTCACCTACAACGGCCCTAGTGGTATCGCAGCTACTCCGGATC 1349
Qy 1021 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1080
Db 1350 CCACAAGCCGTCGTGGACATGGTGGCGGGGGCCCACTGGGGTGTCCTAGCGGGCCTTGCC 1409
Qy 1081 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1140
Db 1410 TACTATTCCATGGTGGGGAACTGGGCTAAGGTCTTGATTGTGATGCTACTCTTTGCTGGC 1469
Qy 1141 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1200
Db 1470 GTTGACGGGCACACCCACGTGACAGGGGGAAGGGTAGCCTCCAGCACCCAGAGCCTCGTG 1529
Qy 1201 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1260
Db 1530 TCCTGGCTCTCACAAGGCCCATCTCAGAAAATCCAACTCGTGAACACCAACGGCAGCTGG 1589
Qy 1261 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1320
Db 1590 CACATCAACAGGACCGCTCTGAATTGCAATGACTCCCTCCAAACTGGGTTCATTGCTGCG 1649
Qy 1321 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1380
Db 1650 CTGTTCTACGCACACAGGTTCAACGCGTCCGGGTGCCCAGAGCGCATGGCTAGCTGCCGC 1709
Qy 1381 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1440
Db 1710 CCCATCGATGAGTTCGCTCAGGGGTGGGGTCCCATCACTCATGATATGCCTGAGAGCTC- 1768
Qy 1441 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1500
Db 1769 -----GGACCAGAGGCCATATTGCTGGCACTACGCGCCTCGACCGTGCGGGATCGTGCCT 1823
Qy 1501 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1560
Db 1824 GCGTCGCAGGTGTGTGGTCCAGTGTATTGCTTCACTCCGAGCCCTGTTGTAGTGGGGACG 1883
Qy 1561 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1620
Db 1884 ACCGATCGTTTCGGCGCTCCTACGTATAGCTGGGGGGAGAATGAGACAGACGTGCTGCTA 1943
Qy 1621 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 1680
Db 1944 CTTAGCAACACGCGGCCGCCTCAAGGCAACTGGTTTGGGTGCACGTGGATGAACAGCACT 2003
Qy 1681 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 1740
Db 2004 GGGTTCACCAAGACGTGCGGGGGCCCTCCGTGCAACATCGGGGGGGTCGGCAACAACA-- 2061
Qy 1741 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 1800
Db 2062 ----CCTTGGTCTGCCCCACGGATTGCTTCCGGAAGCACCCCGAGGCCACTTACACAAAG 2117
Qy 1801 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 1860
Db 2118 TGTGGCTCGGGGCCCTGGTTGACACCCAGGTGCATGGTTGACTACCCATACAGGCTCTGG 2177
Qy 1861 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 1920
Db 2178 CACTACCCCTGCACTGTTAACTTTACCGTCTTTAAGGTCAGGATGTATGTGGGGGGCGTG 2237
Qy 1921 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 1980
Db 2238 GAGCACAGGCTCAATGCTGCATGCAATTGGACTCGAGGAGAGCGCTGTGACTTGGAGGAC 2297
Qy 1981 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2040
Db 2298 AGGGATAGGTCAGAACTCAGCCCGCTGCTGCTGTCTACAACAGAGTGGCAGATACTGCCC 2357
Qy 2041 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2100
Db 2358 TGTTCCTTCACCACCCTACCGGCCCTGTCCACTGGCTTGATCCATCTTCACCGGAACATC 2417
Qy 2101 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2160
Db 2418 GTGGACGTGCAATACCTGTACGGTATAGGGTCGGCAGTTGTCTCCTTTGCAATCAAATGG 2477
Qy 2161 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2220
Db 2478 GAGTATATCCTGTTGCTTTTCCTTCTTCTGGCGGACGCGCGCGTCTGTGCCTGCTTGTGG 2537


Qy 2221 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2280
Db 2538 ATGATGCTGCTGATAGCCCAGGCTGAGGCCACCTTAGAGAACCTGGTGGTCCTCAATGCG 2597
Qy 2281 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATATTTTGTCATCTTTTTCGTGGCTGCTTGG 2340
Db 2598 GCGTCTGTGGCCGGAGCGCATGGCCTTCTCTCCTTCCTCGTGTTCTTCTGCGCCGCCTGG 2657
Qy 2341 TACATCAAGGGTCGGGTAGTCCCCTTAGCTACCTATTCCCTCACTGGCCTGTGGTCCTTT 2400
Db 2658 TACATCAAAGGCAGGCTGGTCCCTGGGGCGGCATATGCTCTCTATGGCGTATGGCCGTTG 2717
Qy 2401 AGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTGCATGGC 2460
Db 2718 CTCCTGCTCTTGCTGGCCTTACCACCACGAGCTTATGCCATGGACCGAGAGATGGCTGCA 2777
Qy 2461 CAGATAGGAGCGGCTCTGCTGGTAATGATCACTCTCTTTACTCTCACCCCCGGGTATAAG 2520
Db 2778 TCGTGCGGAGGCGCGGTTTTTGTAGGTCTGGTACTCTTGACCTTGTCACCATACTATAAG 2837
Qy 2521 ACCCTTCTCAGCCGGTTTTTGTGGTGGTTGTGCTATCTTCTGACCCTGGGGGAAGCTATG 2580
Db 2838 GTGTTCCTCGCTAGGCTCATATGGTGGTTACAATATTTTATCACCAGAGCCGAGGCGCAC 2897
Qy 2581 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGCATCATATGGGCC 2640
Db 2898 TTGCAAGTGTGGGTCCCCCCTCTCAATGTTCGGGGAGGCCGCGATGCCATCATCCTCCTT 2957
Qy 2641 GTCGCCATATTCTACCCAGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 2700
Db 2958 ACATGCGCGGTCCATCCAGAGCTAATCTTTGACATCACCAAACTCCTGCTCGCCATACTC 3017
Qy 2701 GGGCCTGCTTACCTCCTAAAAGGTGCTTTGACGCGCGTGCCGTACTTCGTCAGGGCTCAC 2760
Db 3018 GGTCCGCTCATGGTGCTCCAGGCTGGCATAACTAGAGTGCCGTACTTTGTACGCGCTCAG 3077
Qy 2761 GCTCTACTGAGGATGTGCACCATGGCAAGGCATCTCGCGGGGGGCAGGTACGTCCAGATG 2820
Db 3078 GGGCTCATCCGTGCATGCATGTTAGTGCGGAAGGTCGCTGGAGGCCACTATGTCCAAATG 3137
Qy 2821 GCGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 2880
Db 3138 GCCTTCATGAAGCTGGCCGCGCTGACAGGTACGTACGTATATGACCATCTTACTCCACTG 3197
Qy 2881 TCGGATTGGGCTGCTAGTGGCCTGCGGGACCTGGCGGTCGCCGTTGAGCCTATCATCTTC 2940
Db 3198 CGGGATTGGGCCCACGCGGGCCTACGAGACCTTGCGGTGGCAGTAGAGCCCGTCGTCTTC 3257
Qy 2941 AGTCCGATGGAGAAGAAAGTCATTGTCTGGGGAGCGGAGACAGCTGCTTGTGGGGACATT 3000
Db 3258 TCTGACATGGAGACTAAACTCATCACCTGGGGGGCAGACACCGCGGCGTGTGGGGACATC 3317
Qy 3001 TTACACGGACTTCCCGTGTCCGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3060
Db 3318 ATCTCGGGTCTACCAGTCTCCGCCCGAAGGGGGAAGGAGATACTTCTAGGACCGGCCGAT 3377
Qy 3061 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3120
Db 3378 AGTTTTGGAGAGCAGGGGTGGCGGCTCCTTGCGCCTATCACGGCCTATTCCCAACAAACG 3437
Qy 3121 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCT 3180
Db 3438 CGGGGCCTGCTTGGCTGTATCATCACTAGCCTCACAGGTCGGGACAAGAACCAGGTCGAT 3497
Qy 3181 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3240
Db 3498 GGGGAGGTTCAGGTGCTCTCCACCGCAACGCAATCTTTCCTGGCGACCTGCGTCAATGGC 3557
Qy 3241 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3300
Db 3558 GTGTGTTGGACCGTCTACCATGGTGCCGGCTCGAAGACCCTGGCCGGCCCGAAGGGTCCA 3617
Qy 3301 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGG 3360
Db 3618 ATCACCCAAATGTACACCAATGTAGACCAGGACCTCGTCGGCTGGCCGGCGCCCCCCGGG 3677
Qy 3361 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3420
Db 3678 GCGCGCTCCATGACACCGTGCACCTGCGGCAGCTCGGACCTTTACTTGGTCACGAGGCAT 3737
Qy 3421 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3480
Db 3738 GCTGATGTCGTTCCGGTGCGCCGGCGGGGCGACAGCAGGGGGAGCCTGCTTTCCCCCAGG 3797
Qy 3481 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3540
Db 3798 CCCATCTCCTACCTGAAGGGCTCCTCGGGTGGACCACTGCTTTGCCCTTCGGGGCACGTT 3857
Qy 3541 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3600
Db 3858 GTAGGCATCTTCCGGGCTGCTGTGTGCACCCGGGGGGTTGCGAAGGCGGTGGACTTCATA 3917
Qy 3601 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 3660
Db 3918 CCCGTTGAGTCTATGGAAACTACCATGCGGTCTCCGGTCTTCACAGACAACTCATCCCCT 3977
Qy 3661 CCTGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 3720
Db 3978 CCGGCCGTACCGCAAACATTCCAAGTGGCACATTTACACGCTCCCACTGGCAGCGGCAAG 4037
Qy 3721 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 3780
Db 4038 AGCACCAAAGTGCCGGCTGCATATGCAGCCCAAGGGTACAAGGTGCTCGTCCTAAACCCG 4097
Qy 3781 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 3840
Db 4098 TCCGTTGCCGCCACATTGGGCTTTGGAGCGTATATGTCCAAGGCACATGGCATCGAGCCT 4157
Qy 3841 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 3900
Db 4158 AACATCAGAACTGGGGTAAGGACCATCACCACGGGCGGCCCCATCACGTACTCCACCTAT 4217
Qy 3901 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 3960
Db 4218 TGCAAGTTCCTTGCCGACGGTGGATGCTCCGGGGGCGCCTATGACATCATAATATGTGAT 4277
Qy 3961 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4020
Db 4278 GAATGCCACTCAACTGACTCGACTACCATCTTGGGCATCGGCACAGTCCTGGATCAGGCA 4337
Qy 4021 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4080
Db 4338 GAGACGGCTGGAGCGCGGCTCGTCGTGCTCGCCACCGCCACGCCTCCGGGATCGATCACC 4397
Qy 4081 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4140
Db 4398 GTGCCACACCCCAACATCGAGGAAGTGGCCCTGTCCAACACTGGAGAGATTCCCTTCTAT 4457
Qy 4141 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4200
Db 4458 GGCAAAGCCATCCCCATTGAGGCCATCAAGGGGGGAAGGCATCTCATCTTCTGCCATTCC 4517
Qy 4201 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4260
Db 4518 AAGAAGAAGTGTGACGAGCTCGCCGCAAAGCTGACAGGCCTCGGACTCAATGCTGTAGCG 4577
Qy 4261 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4320
Db 4578 TATTACCGGGGTCTCGATGTGTCCGTCATACCGACTAGCGGAGACGTCGTTGTCGTGGCA 4637
Qy 4321 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4380
Db 4638 ACAGACGCTCTAATGACGGGTTTTACCGGCGACTTTGACTCAGTGATCGACTGCAACACA 4697
Qy 4381 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4440


Db 4698 TGTGTCACCCAGACAGTCGATTTCAGCTTGGATCCCACCTTCACCATTGAGACGACAACG 4757
Qy 4441 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4500
Db 4758 CTGCCCCAAGACGCGGTGTCGCGTGCGCAGCGGCGAGGTAGGACTGGCAGGGGCAGGAGT 4817
Qy 4501 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4560
Db 4818 GGCATCTACAGGTTTGTGACTCCAGGAGAACGGCCCTCAGGCATGTTCGACTCCTCGGTC 4877
Qy 4561 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4620
Db 4878 CTGTGTGAGTGCTATGACGCAGGCTGCGCTTGGTATGAGCTCACGCCCGCTGAGACCTCG 4937
Qy 4621 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 4680
Db 4938 GTTAGGTTGCGGGCTTACCTAAATACACCAGGGTTGCCCGTCTGCCAGGACCACCTAGAG 4997
Qy 4681 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 4740
Db 4998 TTCTGGGAGAGCGTCTTCACAGGCCTCACCCACATAGATGCCCACTTCTTGTCCCAGACC 5057
Qy 4741 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 4800
Db 5058 AAACAGGCAGGAGACAACCTCCCCTACCTGGTAGCATACCAAGCCACAGTGTGCGCCAGG 5117
Qy 4801 GCCAAAGCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 4860
Db 5118 GCTCAGGCTCCACCTCCATCGTGGGACCAAATGTGGAAGTGTCTCATACGGCTAAAGCCC 5177
Qy 4861 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 4920
Db 5178 ACACTGCATGGGCCAACGCCCCTGCTGTACAGGCTAGGAGCCGTTCAAAATGAGGTCACT 5237
Qy 4921 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 4980
Db 5238 CTCACACACCCCATAACCAAATACATCATGGCATGCATGTCGGCTGACCTGGAGGTCGTC 5297
Qy 4981 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5040
Db 5298 ACTAGCACCTGGGTGCTAGTAGGCGGAGTCCTTGCGGCTCTGGCCGCGTACTGCCTGACG 5357
Qy 5041 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5100
Db 5358 ACAGGCAGCGTGGTCATTGTGGGCAGGATCATCTTGTCCGGGAGGCCAGCTGTTATTCCC 5417
Qy 5101 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5160
Db 5418 GACAGGGAAGTCCTCTACCAGGAGTTCGATGAGATGGAAGAGTGTGCTTCACACCTCCCT 5477
Qy 5161 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5220
Db 5478 TACATCGAGCAAGGAATGCAGCTCGCCGAGCAATTCAAACAGAAGGCGCTCGGATTGCTG 5537
Qy 5221 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5280
Db 5538 CAAACAGCCACCAAGCAAGCGGAGGCTGCTGCTCCCGTGGTGGAGTCCAAGTGGCGAGCC 5597
Qy 5281 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5340
Db 5598 CTTGAGGTCTTCTGGGCGAAACACATGTGGAACTTCATCAGCGGGATACAGTACTTGGCA 5657
Qy 5341 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5400
Db 5658 GGCCTATCCACTCTGCCTGGAAACCCCGCGATAGCATCATTGATGGCTTTTACAGCCTCT 5717
Qy 5401 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5460
Db 5718 ATCACCAGCCCGCTCACCACCCAAAATACCCTCCTGTTTAACATCTTGGGGGGATGGGTG 5777
Qy 5461 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5520
Db 5778 GCTGCCCAACTCGCTCCCCCCAGCGCTGCTTCGGCTTTCGTGGGCGCCGGCATTGCCGGT 5837
Qy 5521 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5580
Db 5838 GCGGCCGTTGGCAGCATAGGTCTCGGGAAGGTACTTGTGGACATTCTGGCGGGCTATGGG 5897
Qy 5581 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5640
Db 5898 GCGGGGGTGGCTGGCGCACTCGTGGCCTTTAAGGTCATGAGCGGCGAGATGCCCTCCACT 5957
Qy 5641 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 5700
Db 5958 GAGGATCTGGTTAATTTACTCCCTGCCATCCTTTCTCCTGGCGCCCTGGTTGTCGGGGTC 6017
Qy 5701 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 5760
Db 6018 GTGTGCGCAGCAATACTGCGTCGGCACGTGGGCCCGGGAGAGGGGGCTGTGCAGTGGATG 6077
Qy 5761 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 5820
Db 6078 AACCGGCTGATAGCGTTCGCTTCGCGGGGTAACCACGTCTCCCCCACGCACTATGTGCCC 6137
Qy 5821 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 5880
Db 6138 GAGAGCGACGCCGCGGCGCGTGTTACTCAGATCCTCTCCAGCCTTACCATCACTCAGTTG 6197
Qy 5881 CTCAGAAGACTCCACAACTGGATTACTGAGGACTGCCCCATCCCATGCGGCGGCTCGTGG 5940
Db 6198 CTGAAGAGGCTTCATCAGTGGATTAATGAGGACTGCTCCACGCCTTGTTCCGGCTCGTGG 6257
Qy 5941 CTCCGCGATGTGTGGGACTGGGTTTGCACCATCCTAACAGACTTTAAAAATTGGCTGACC 6000
Db 6258 CTAAAGGATGTTTGGGACTGGATATGCACGGTGTTGAGTGACTTCAAGACTTGGCTCCAG 6317
Qy 6001 TCCAAATTATTCCCAAAGATGCCCGGCCTCCCCTTTGTCTCCTGTCAAAAGGGGTACAAG 6060
Db 6318 TCCAAGCTCCTGCCGCGGTTACCGGGACTCCCTTTCCTGTCATGCCAACGCGGGTACAAG 6377
Qy 6061 GGCGTGTGGGCCGGCACTGGCATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6120
Db 6378 GGAGTCTGGCGGGGGGATGGCATCATGCAAACCACCTGCCCATGTGGAGCACAGATCACC 6437
Qy 6121 GGCAATGTCCGCTTGGGCTCCATGAGAATCACGGGGCCTAAGACCTGCATGAATATCTGG 6180
Db 6438 GGACATGTCAAAAATGGCTCCATGAGGATTGTTGGGCCAAAAACCTGCAGCAACACGTGG 6497
Qy 6181 CAGGGGACCTTTCCTATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCGCCA 6240
Db 6498 CATGGAACATTCCCCATCAACGCATACACCACGGGCCCCTGCACGCCCTCCCCAGCGCCG 6557
Qy 6241 AACTTTAAGGTCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6300
Db 6558 AACTATTCCAGGGCGCTGTGGCGGGTGGCTGCTGAGGAGTACGTGGAGGTTACGCGGGTG 6617
Qy 6301 GGGTCATACCACTACATAACAGGACTCACCACTGATAACTTGAAAGTCCCCTGCCAACTA 6360
Db 6618 GGGGATTTCCACTACGTGACGGGCATGACCACTGACAACGTGAAATGCCCATGCCAGGTT 6677
Qy 6361 CCCTCTCCCGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCACA 6420
Db 6678 CCAGCCCCTGAATTTTTCACGGAGGTGGATGGAGTACGGTTGCACAGGTATGCTCCAGTG 6737
Qy 6421 CCGAAGCCGTTTTTCCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6480
Db 6738 TGCAAACCTCTCCTACGAGAGGAGGTCGTATTCCAGGTCGGGCTCAACCAGTACCTGGTC 6797
Qy 6481 GGGTCCCAGCTTCCTTGCGACCCTGAACCCGACACAGACGTATTGATGTCCATGCTAACA 6540
Db 6798 GGGTCACAGCTCCCATGTGAGCCCGAACCGGATGTGGCAGTGCTCACTTCCATGCTCACC 6857
Qy 6541 GATCCATCTCATATCACGGCGGAGACTGCAGCGCGGCGTTTAGCGCGGGGGTCACCCCCA 6600
Db 6858 GACCCCTCTCATATTACAGCAGAGACGGCCAAGCGTAGGCTGGCCAGGGGGTCTCCCCCC 6917


Qy 6601 TCCGAGGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 6660
Db 6918 TCCTTGGCCAGCTCTTCAGCTAGCCAGTTGTCTGCGCCTTCTTTGAAGGCGACATGTACT 6977
Qy 6661 ACCCACGGCAAAGCCTATGATGTGGACATGGTGGATGCTAACCT----------GTTC 6708
Db 6978 ACCCATCATGACTCCCCGGACGCTGACCTCATCGAGGCCAACCTCCTGTGGCGGCAGGAG 7037
Qy 6709 ATGGGGGGCGATGTGACTCGGATAGAGTCTGGGTCCAAAGTGGTCGTTCTGGACTCTCTC 6768
Db 7038 ATGGGCGGGAACATCACCCGTGTGGAGTCAGAAAATAAGGTGGTAATCCTGGACTCTTTC 7097
Qy 6769 GACCCAATGGTCGAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCAGAATACATGCTC 6828
Db 7098 GATCCGATTCGGGCGGTGGAGGATGAGAGGGAAATATCCGTCCCGGCGGAGATCCTGCGA 7157
Qy 6829 CCCAAGAAGAGGTTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCG 6888
Db 7158 AAACCCAGGAAGTTCCCCCCAGCGTTGCCCATATGGGCACGCCCGGATTACAACCCTCCA 7217
Qy 6889 CTTGTGGAATCGTGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTC 6948
Db 7218 CTGCTAGAGTCCTGGAAGGACCCGGACTACGTCCCCCCGGTGGTACACGGGTGCCCTTTG 7277
Qy 6949 CCTCCTCCTAGGAAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGT 7008
Db 7278 CCATCTACCAAGGCCCCCCCAATACCACCTCCACGGAGGAAGAGGACGGTTGTCCTGACA 7337
Qy 7009 GAGGACTCCATAGGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCC 7068
Db 7338 GAGTCCACCGTGTCTTCTGCCTTGGCGGAGCTCGCTACTAAGACCTTTGGCAGCTCCGGG 7397
Qy 7069 CCAAGCGGCGATTCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACG 7128
Db 7398 TCGTCGGCCGTTGACAGCGG---CACGGCGACTGGCCCTCCCGA---------TCAGGCC 7445
Qy 7129 CCTCCTGATGAGTTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCTCGAG 7188
Db 7446 TCCGACGACGGCGACAAAGGATCCGACGTTGAGTCGTACTCCTCCATGCCCCCCCTCGAG 7505
Qy 7189 GGGGAGCTTGGAGATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCAG 7248
Db 7506 GGAGAGCCAGGGGACCCCGACCT---------- 7528
Qy 7249 GGGGGGGTGGCAGCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGAC 7308
Db 7529 ----------CAGCGACGGGTCTTGGTCTACCGTGAGCGGGGAAGCTGGT 7568
Qy 7309 GACTCCGTCGTGTGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGT 7368
Db 7569 GAGGACGTCGTCTGCTGCTCAATGTCCTATACATGGACAGGTGCCTTGATCACGCCATGC 7628
Qy 7369 AGTCCCGAAGAGGAGAAGTTACCGATTAACCCCTTGAGCAACTCCCTGTTGCGATATCAC 7428
Db 7629 GCTGCGGAGGAGAGCAAGTTGCCCATCAATCCGTTGAGCAACTCTTTGCTGCGTCACCAC 7688
Qy 7429 AACAAGGTGTACTGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTT 7488
Db 7689 AGTATGGTCTACTCCACAACATCTCGCAGCGCAAGTCTGCGGCAGAAGAAGGTCACCTTT 7748
Qy 7489 GATAGGATGCAAGTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCG 7548
Db 7749 GACAGACTGCAAGTCCTGGACGACCACTACCGGGACGTGCTCAAGGAGATGAAGGCGAAG 7808
Qy 7549 GCCTCCAAGGTCACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCC 7608
Db 7809 GCGTCCACAGTTAAGGCTAGGCTTCTATCTATAGAGGAGGCCTGCAAACTGACGCCCCCA 7868
Qy 7609 CATTCTGCAAGATCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGG 7668
Db 7869 CATTCGGCCAAATCCAAATTTGGCTACGGGGCGAAGGACGTCCGGAGCCTATCCAGCAGG 7928
Qy 7669 GCCGTTAACCACATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATT 7728
Db 7929 GCCGTCAACCACATCCGCTCCGTGTGGGAGGACTTGCTGGAAGACACTGAAACACCAATT 7988
Qy 7729 CCCACAACCATTATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAG 7788
Db 7989 GATACCACCATCATGGCAAAAAATGAGGTTTTCTGCGTCCAACCAGAGAAAGGAGGCCGC 8048
Qy 7789 AAAGCAGCTCGCCTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCC 7848
Db 8049 AAGCCAGCTCGCCTTATCGTATTCCCAGACCTGGGGGTACGTGTATGCGAGAAGATGGCC 8108
Qy 7849 CTTTATGACATTACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAG 7908
Db 8109 CTTTACGACGTGGTCTCCACCCTTCCTCAGGCCGTGATGGGCCCCTCATACGGATTCCAG 8168
Qy 7909 TATTCCCCCGCTCAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCT 7968
Db 8169 TACTCTCCTGGGCAGCGGGTCGAGTTCCTGGTGAATACCTGGAAATCAAAGAAATGCCCT 8228
Qy 7969 ATGGGTTTTTCGTATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGG 8028
Db 8229 ATGGGCTTCTCATATGACACCCGCTGCTTTGACTCAACGGTCACTGAGAATGACATCCGT 8288
Qy 8029 ACTGAGGAGTCCATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACAC 8088
Db 8289 ACTGAGGAATCAATTTACCAATGTTGTGACTTGGCCCCCGAAGCCAGGCAGGCCATAAGG 8348
Qy 8089 TCGCTAACTGAGAGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGC 8148
Db 8349 TCGCTCACAGAGCGGCTTTATGTCGGGGGTCCCCTGACTAATTCGAAGGGGCAGAACTGC 8408
Qy 8149 GGGTACAGGCGTTGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACA 8208
Db 8409 GGTTATCGCCGGTGCCGCGCAAGTGGCGTGCTGACGACTAGCTGCGGCAACACCCTCACA 8468
Qy 8209 TGCTACGTGAAAGCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTG 8268
Db 8469 TGTTACTTGAAGGCCACTGCGGCCTGTCGAGCTGCAAAGCTCCAGGACTGCACGATGCTC 8528
Qy 8269 GTATGCGGCGATGACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGG 8328
Db 8529 GTGAACGGAGACGACCTTGTCGTTATCTGTGAGAGTGCGGGAACCCAGGAGGATGCGGCG 8588
Qy 8329 AACCTGAGAGCCTTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCC 8388
Db 8589 GCCCTACGAGCCTTCACGGAGGCTATGACTAGGTATTCCGCCCCCCCCGGGGACCCGCCC 8648
Qy 8389 AGACCGGAGTATGATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTG 8448
Db 8649 CAACCAGAATACGACTTGGAGCTGATAACGTCATGCTCCTCCAATGTGTCGGTCGCGCAC 8708
Qy 8449 GGCCCACAAGGCCGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGG 8508
Db 8709 GATGCATCCGGCAAAAGGGTGTACTACCTCACCCGTGACCCCACCACCCCCCTCGCACGG 8768
Qy 8509 GCTGCCTGGGAAACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAG 8568
Db 8769 GCTGCGTGGGAGACAGTTAGACACACTCCAGTCAACTCCTGGCTAGGCAATATCATCATG 8828
Qy 8569 TACGCCCCGACCATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATG 8628
Db 8829 TATGCGCCCACCCTATGGGCGAGGATGATTCTGATGACTCATTTCTTCTCTATCCTTCTA 8888
Qy 8629 GCTCAAGACACGCTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTG 8688
Db 8889 GCTCAGGAGCAACTTGAAAAAGCCCTGGATTGTCAGATCTACGGGGCCTGTTACTCCATT 8948
Qy 8689 AGTCCCTTGGACCTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTG 8748
Db 8949 GAGCCACTTGACCTACCTCAGATCATTGAACGACTCCATGGTCTTAGCGCATTTTCACTC 9008
Qy 8749 CACACATACACTCCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCG 8808

```
Db       9009 CACAGTTACTCTCCAGGTGAGATCAATAGGGTGGCTTCATGCCTCAGGAAACTTGGGGTA 9068

Qy       8809 CCACCCCTCAGAGCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGT 8868

Db       9069 CCGCCTTTGCGAGTCTGGAGACATCGGGCCAGAAGTGTCCGCGCTAAGCTACTGTCCCAG 9128

Qy       8869 GGGGGGAGAGCGGCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTC 8928

Db       9129 GGGGGGAGGGCTGCCACTTGCGGCAAGTACCTCTTCAACTGGGCAGTAAAGACCAAGCTT 9188

Qy       8929 AAACTCACTCCATTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGC 8988

Db       9189 AAACTCACTCCAATCCCGGCTGCGTCCCAGCTAGACTTGTCCGGCTGGTTCGTTGCTGGT 9248

Qy       8989 GCCGGCGGGGGCGACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTT 9048

Db       9249 TACAACGGGGGAGACATATATCACAGCCTGTCTCGTGCCCGACCCCGTTGGTTCATGTTG 9308

Qy       9049 GGCCTACTCCTACTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCGG 9099

Db       9309 TGCCTACTCCTACTTTCTGTAGGGGTAGGCATCTACCTGCTCCCCAACCGG 9359
```

RESULT 14
ADF88596
ID ADF88596 standard; DNA; 9413 BP.
XX
AC ADF88596;
XX
DT 26-FEB-2004 (first entry)
XX
DE Hepatitis C virus NS3 gene, SEQ ID No 5.
XX
KW liver cancer; hepatitis-C virus; NS3 gene; carcinogenicity; ds; gene.
XX
OS Hepatitis C virus.
XX
PN JP2003210181-A.
XX
PD 29-JUL-2003.
XX
PF 30-MAY-2002; 2002JP-00158335.
XX
PR 16-NOV-2001; 2001JP-00352443.
XX
PA (SHIN-) ZH SHINSANGYO SOZO KENKYU KIKO.
XX
DR WPI; 2003-819836/77.
DR P-PSDB; ADF88597.
XX
PT Diagnosing liver cancer, involves amplifying amino terminal region of
PT hepatitis-C virus gene using predetermined primer and determining
PT hepatitis-C virus in base sequence of obtained DNA fragment.
XX
PS Disclosure; SEQ ID NO 5; 36pp; Japanese.
XX
CC The invention relates to the novel testing method for diagnosing liver
CC cancer. The novel method comprises amplifying the amino terminal region
CC of a hepatitis-C virus NS3 gene using a predetermined primer and
CC determining the hepatitis-C virus in a base sequence of the obtained DNA
CC fragment. The novel testing method is useful for diagnosing liver cancer
CC and also used in a gene amplification technique, a clinical laboratory
CC test reagent, a polymerase chain reaction, a base sequence analysis and
CC genetic engineering. The method enables the detection of a hepatitis-C
CC virus having high carcinogenicity with high specificity. This
CC polynucleotide sequence represents the hepatitis-C virus NS3 gene of the
CC invention.
XX
SQ Sequence 9413 BP; 1886 A; 2816 C; 2694 G; 2017 T; 0 U; 0 Other;

Query Match 45.5%; Score 4141.8; DB 10; Length 9413;
Best Local Similarity 66.6%; Pred. No. 0;

Matches 6066; Conservative 0; Mismatches 2952; Indels 93; Gaps 6;

```
Qy          1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 60

Db        330 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCCGCCCACAG 389

Qy         61 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 120

Db        390 GACGTTAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGG 449

Qy        121 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 180

Db        450 GGCCCCAGGTTGGGTGTGCGCGCGACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTGGA 509

Qy        181 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 240

Db        510 AGGCGACAACCTATCCCCAAGGCTCGCCGGCCCGAGGGTAGGACCTGGGCTCAGCCCGGG 569

Qy        241 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 300

Db        570 TACCCTTGGCCCCTCTATGGCAACGAGGGTATGGGGTGGGCAGGATGGCTCCTGTCACCC 629

Qy        301 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 360

Db        630 CGTGGCTCTCGGCCTAGTTGGGGCCCCACAGACCCCCGGCGTAGGTCGCGTAATTTGGGT 689

Qy        361 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 420

Db        690 AAGGTCATCGATACCCTTACATGCGGCTTCGCCGACCTCATGGGGTACATTCCGCTTGTC 749

Qy        421 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 480

Db        750 GGCGCCCCCCTAGGGGGCGCTGCCAGGGCCCTGGCACATGGTGTCCGGGTTCTGGAGGAC 809

Qy        481 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 540

Db        810 GGCGTGAACTATGCAACAGGGAATCTGCCCGGTTGCTCTTTCTCTATCTTCCTCTTAGCT 869

Qy        541 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 600

Db        870 TTGCTGTCTTGTTTGACCATCCCAGCTTCCGCTTACGAGGTGCGCAACGTGTCCGGGATA 929

Qy        601 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 660

Db        930 TACCATGTCACGAACGACTGCTCCAACTCAAGTATTGTGTATGAGGCAGCGGACATGATC 989

Qy        661 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 720

Db        990 ATGCACACCCCCGGGTGCGTGCCCTGCGTCCGGGAGAGTAATTTCTCCCGTTGCTGGGTA 1049

Qy        721 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 780

Db       1050 GCGCTCACTCCCACGCTCGCGGCCAGGAACAGCAGCATCCCCACCACGACAATACGACGC 1109

Qy        781 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 840

Db       1110 CACGTCGATTTGCTCGTTGGGGCGGCTGCTCTCTGTTCCGCTATGTACGTTGGGGATCTC 1169

Qy        841 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 900

Db       1170 TGCGGATCCGTTTTTCTCGTCTCCCAGCTGTTCACCTTCTCACCTCGCCGGTATGAGACG 1229

Qy        901 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 960

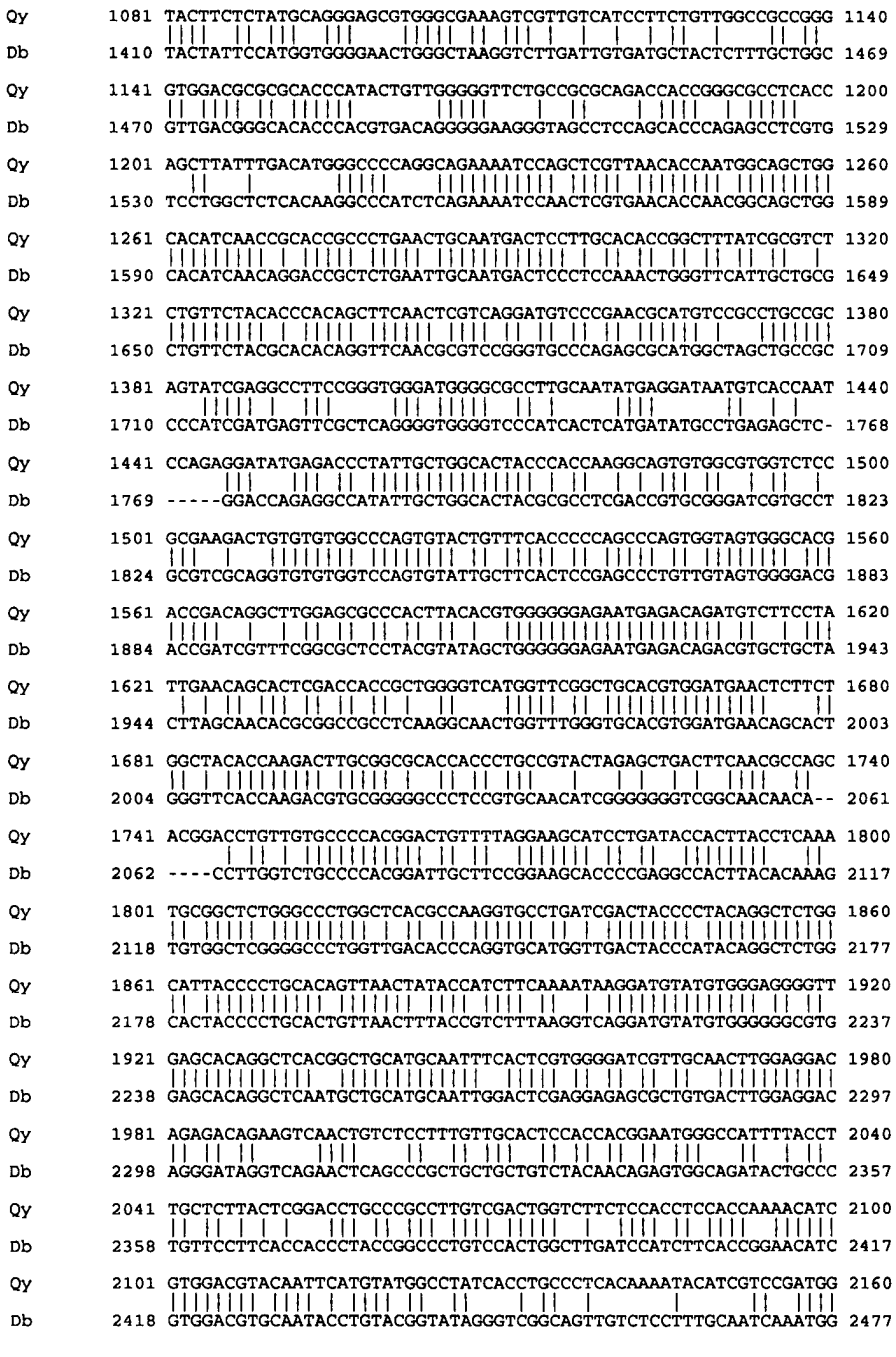
Db       1230 GTACAAGATTGCAATTGCTCAATCTATCCCGGCCACGTATCAGGTCACCGCATGGCTTGG 1289

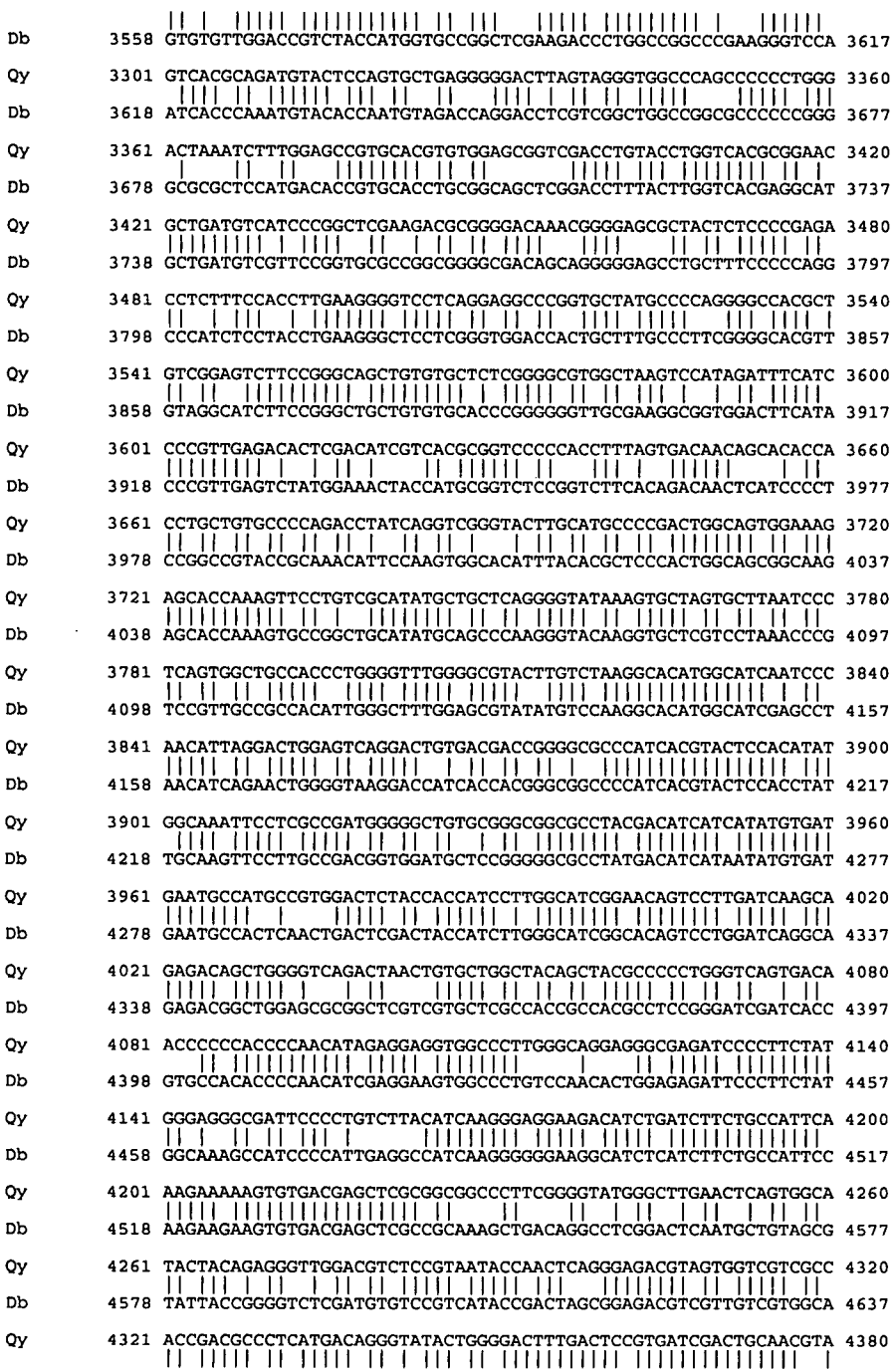
Qy        961 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1020

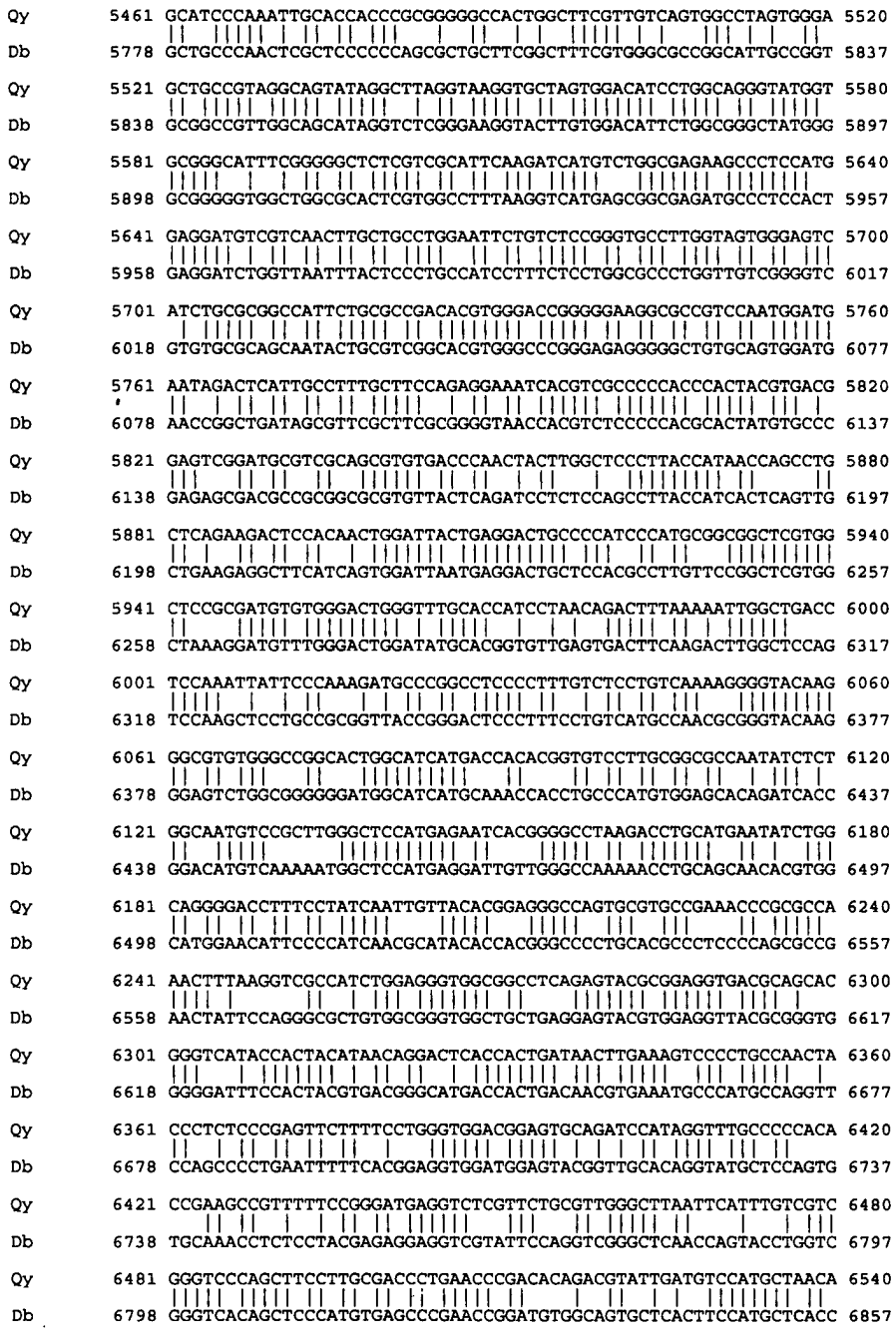
Db       1290 GATATGATGATGAACTGGTCACCTACAACGGCCCTAGTGGTATCGCAGCTACTCCGGATC 1349

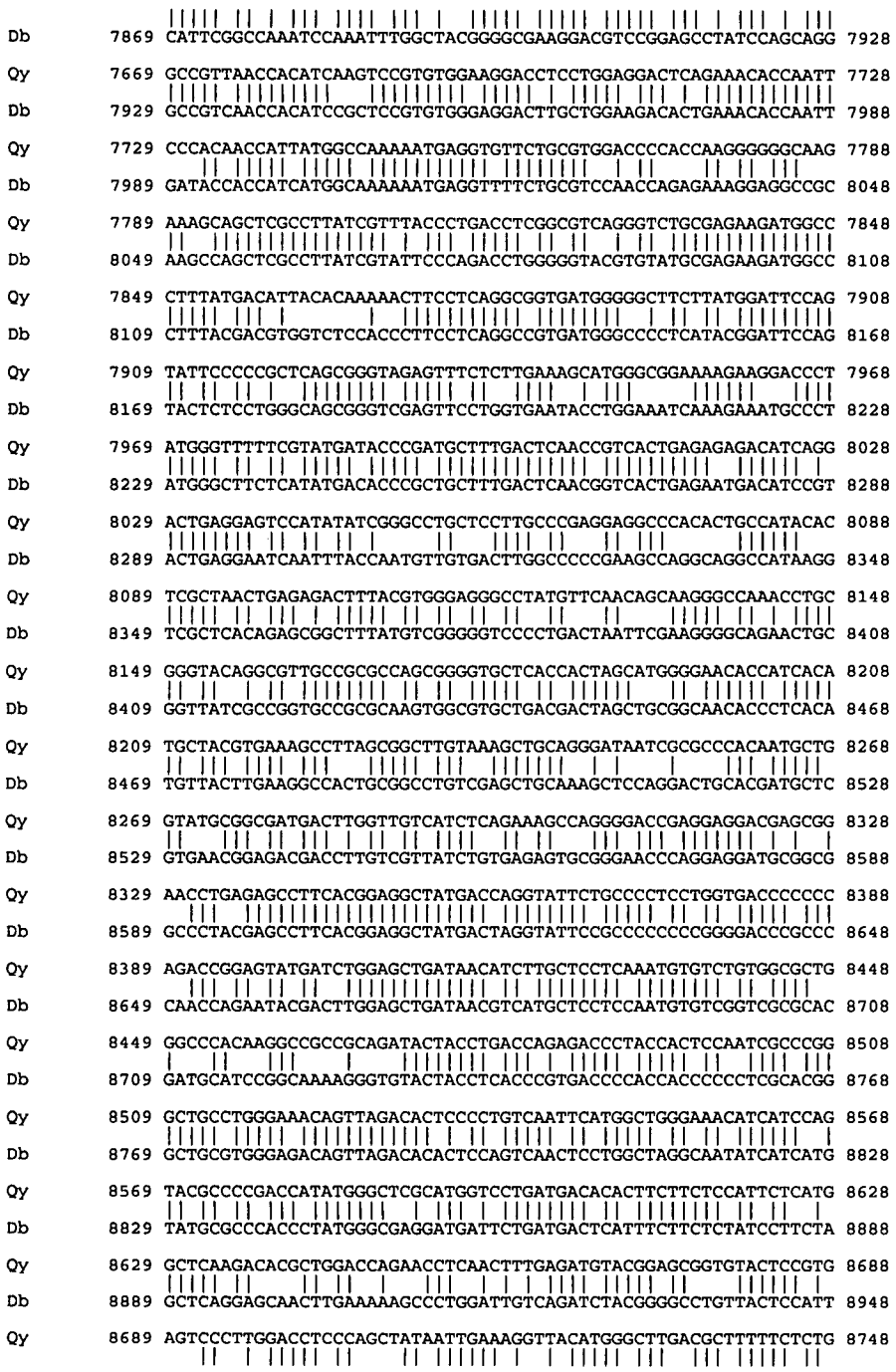
Qy       1021 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1080

Db       1350 CCACAAGCCGTCGTGGACATGGTGGCGGGGGCCCACTGGGGTGTCCTAGCGGGCCTTGCC 1409
```


Qy 1081 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1140
Db 1410 TACTATTCCATGGTGGGGAACTGGGCTAAGGTCTTGATTGTGATGCTACTCTTTGCTGGC 1469
Qy 1141 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1200
Db 1470 GTTGACGGGCACACCCACGTGACAGGGGGAAGGGTAGCCTCCAGCACCCAGAGCCTCGTG 1529
Qy 1201 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1260
Db 1530 TCCTGGCTCTCACAAGGCCCATCTCAGAAAATCCAACTCGTGAACACCAACGGCAGCTGG 1589
Qy 1261 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1320
Db 1590 CACATCAACAGGACCGCTCTGAATTGCAATGACTCCCTCCAAACTGGGTTCATTGCTGCG 1649
Qy 1321 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1380
Db 1650 CTGTTCTACGCACACAGGTTCAACGCGTCCGGGTGCCCAGAGCGCATGGCTAGCTGCCGC 1709
Qy 1381 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1440
Db 1710 CCCATCGATGAGTTCGCTCAGGGGTGGGGTCCCATCACTCATGATATGCCTGAGAGCTC- 1768
Qy 1441 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1500
Db 1769 -----GGACCAGAGGCCATATTGCTGGCACTACGCGCCTCGACCGTGCGGGATCGTGCCT 1823
Qy 1501 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1560
Db 1824 GCGTCGCAGGTGTGTGGTCCAGTGTATTGCTTCACTCCGAGCCCTGTTGTAGTGGGGACG 1883
Qy 1561 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1620
Db 1884 ACCGATCGTTTCGGCGCTCCTACGTATAGCTGGGGGGAGAATGAGACAGACGTGCTGCTA 1943
Qy 1621 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 1680
Db 1944 CTTAGCAACACGCGGCCGCCTCAAGGCAACTGGTTTGGGTGCACGTGGATGAACAGCACT 2003
Qy 1681 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 1740
Db 2004 GGGTTCACCAAGACGTGCGGGGGCCCTCCGTGCAACATCGGGGGGGTCGGCAACAACA-- 2061
Qy 1741 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 1800
Db 2062 ----CCTTGGTCTGCCCCACGGATTGCTTCCGGAAGCACCCCGAGGCCACTTACACAAAG 2117
Qy 1801 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 1860
Db 2118 TGTGGCTCGGGGCCCTGGTTGACACCCAGGTGCATGGTTGACTACCCATACAGGCTCTGG 2177
Qy 1861 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 1920
Db 2178 CACTACCCCTGCACTGTTAACTTTACCGTCTTTAAGGTCAGGATGTATGTGGGGGGCGTG 2237
Qy 1921 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 1980
Db 2238 GAGCACAGGCTCAATGCTGCATGCAATTGGACTCGAGGAGAGCGCTGTGACTTGGAGGAC 2297
Qy 1981 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2040
Db 2298 AGGGATAGGTCAGAACTCAGCCCGCTGCTGCTGTCTACAACAGAGTGGCAGATACTGCCC 2357
Qy 2041 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2100
Db 2358 TGTTCCTTCACCACCCTACCGGCCCTGTCCACTGGCTTGATCCATCTTCACCGGAACATC 2417
Qy 2101 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2160
Db 2418 GTGGACGTGCAATACCTGTACGGTATAGGGTCGGCAGTTGTCTCCTTTGCAATCAAATGG 2477
Qy 2161 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2220
Db 2478 GAGTATATCCTGTTGCTTTTCCTTCTTCTGGCGGACGCGCGCGTCTGTGCCTGCTTGTGG 2537
Qy 2221 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2280
Db 2538 ATGATGCTGCTGATAGCCCAGGCTGAGGCCACCTTAGAGAACCTGGTGGTCCTCAATGCG 2597
Qy 2281 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATATTTTGTCATCTTTTTCGTGGCTGCTTGG 2340
Db 2598 GCGTCTGTGGCCGGAGCGCATGGCCTTCTCTCCTTCCTCGTGTTCTTCTGCGCCGCCTGG 2657
Qy 2341 TACATCAAGGGTCGGGTAGTCCCCTTAGCTACCTATTCCCTCACTGGCCTGTGGTCCTTT 2400
Db 2658 TACATCAAAGGCAGGCTGGTCCCTGGGGCGGCATATGCTCTCTATGGCGTATGGCCGTTG 2717
Qy 2401 AGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTGCATGGC 2460
Db 2718 CTCCTGCTCTTGCTGGCCTTACCACCACGAGCTTATGCCATGGACCGAGAGATGGCTGCA 2777
Qy 2461 CAGATAGGAGCGGCTCTGCTGGTAATGATCACTCTCTTTACTCTCACCCCCGGGTATAAG 2520
Db 2778 TCGTGCGGAGGCGCGGTTTTTGTAGGTCTGGTACTCTTGACCTTGTCACCATACTATAAG 2837
Qy 2521 ACCCTTCTCAGCCGGTTTTTGTGGTGGTTGTGCTATCTTCTGACCCTGGGGGAAGCTATG 2580
Db 2838 GTGTTCCTCGCTAGGCTCATATGGTGGTTACAATATTTTATCACCAGAGCCGAGGCGCAC 2897
Qy 2581 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGCATCATATGGGCC 2640
Db 2898 TTGCAAGTGTGGGTCCCCCCTCTCAATGTTCGGGGAGGCCGCGATGCCATCATCCTCCTT 2957
Qy 2641 GTCGCCATATTCTACCCAGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 2700
Db 2958 ACATGCGCGGTCCATCCAGAGCTAATCTTTGACATCACCAAACTCCTGCTCGCCATACTC 3017
Qy 2701 GGGCCTGCTTACCTCCTAAAAGGTGCTTTGACGCGCGTGCCGTACTTCGTCAGGGCTCAC 2760
Db 3018 GGTCCGCTCATGGTGCTCCAGGCTGGCATAACTAGAGTGCCGTACTTTGTACGCGCTCAG 3077
Qy 2761 GCTCTACTGAGGATGTGCACCATGGCAAGGCATCTCGCGGGGGGCAGGTACGTCCAGATG 2820
Db 3078 GGGCTCATCCGTGCATGCATGTTAGTGCGGAAGGTCGCTGGAGGCCACTATGTCCAAATG 3137
Qy 2821 GCGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 2880
Db 3138 GCCTTCATGAAGCTGGCCGCGCTGACAGGTACGTACGTATATGACCATCTTACTCCACTG 3197
Qy 2881 TCGGATTGGGCTGCTAGTGGCCTGCGGGACCTGGCGGTCGCCGTTGAGCCTATCATCTTC 2940
Db 3198 CGGGATTGGGCCCACGCGGGCCTACGAGACCTTGCGGTGGCAGTAGAGCCCGTCGTCTTC 3257
Qy 2941 AGTCCGATGGAGAAGAAAGTCATTGTCTGGGGAGCGGAGACAGCTGCTTGTGGGGACATT 3000
Db 3258 TCTGACATGGAGACTAAACTCATCACCTGGGGGGCAGACACCGCGGCGTGTGGGGACATC 3317
Qy 3001 TTACACGGACTTCCCGTGTCCGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3060
Db 3318 ATCTCGGGTCTACCAGTCTCCGCCCGAAGGGGGAAGGAGATACTTCTAGGACCGGCCGAT 3377
Qy 3061 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3120
Db 3378 AGTTTTGGAGAGCAGGGGTGGCGGCTCCTTGCGCCTATCACGGCCTATTCCCAACAAACG 3437
Qy 3121 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCT 3180
Db 3438 CGGGGCCTGCTTGGCTGTATCATCACTAGCCTCACAGGTCGGGACAAGAACCAGGTCGAT 3497
Qy 3181 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3240
Db 3498 GGGGAGGTTCAGGTGCTCTCCACCGCAACGCAATCTTTCCTGGCGACCTGCGTCAATGGC 3557
Qy 3241 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3300


Db 3558 GTGTGTTGGACCGTCTACCATGGTGCCGGCTCGAAGACCCTGGCCGGCCCGAAGGGTCCA 3617
Qy 3301 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGG 3360
Db 3618 ATCACCCAAATGTACACCAATGTAGACCAGGACCTCGTCGGCTGGCCGGCGCCCCCCGGG 3677
Qy 3361 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3420
Db 3678 GCGCGCTCCATGACACCGTGCACCTGCGGCAGCTCGGACCTTTACTTGGTCACGAGGCAT 3737
Qy 3421 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3480
Db 3738 GCTGATGTCGTTCCGGTGCGCCGGCGGGGCGACAGCAGGGGGAGCCTGCTTTCCCCCAGG 3797
Qy 3481 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3540
Db 3798 CCCATCTCCTACCTGAAGGGCTCCTCGGGTGGACCACTGCTTTGCCCTTCGGGGCACGTT 3857
Qy 3541 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3600
Db 3858 GTAGGCATCTTCCGGGCTGCTGTGTGCACCCGGGGGGTTGCGAAGGCGGTGGACTTCATA 3917
Qy 3601 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 3660
Db 3918 CCCGTTGAGTCTATGGAAACTACCATGCGGTCTCCGGTCTTCACAGACAACTCATCCCCT 3977
Qy 3661 CCTGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 3720
Db 3978 CCGGCCGTACCGCAAACATTCCAAGTGGCACATTTACACGCTCCCACTGGCAGCGGCAAG 4037
Qy 3721 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 3780
Db 4038 AGCACCAAAGTGCCGGCTGCATATGCAGCCCAAGGGTACAAGGTGCTCGTCCTAAACCCG 4097
Qy 3781 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 3840
Db 4098 TCCGTTGCCGCCACATTGGGCTTTGGAGCGTATATGTCCAAGGCACATGGCATCGAGCCT 4157
Qy 3841 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 3900
Db 4158 AACATCAGAACTGGGGTAAGGACCATCACCACGGGCGGCCCCATCACGTACTCCACCTAT 4217
Qy 3901 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 3960
Db 4218 TGCAAGTTCCTTGCCGACGGTGGATGCTCCGGGGGCGCCTATGACATCATAATATGTGAT 4277
Qy 3961 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4020
Db 4278 GAATGCCACTCAACTGACTCGACTACCATCTTGGGCATCGGCACAGTCCTGGATCAGGCA 4337
Qy 4021 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4080
Db 4338 GAGACGGCTGGAGCGCGGCTCGTCGTGCTCGCCACCGCCACGCCTCCGGGATCGATCACC 4397
Qy 4081 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4140
Db 4398 GTGCCACACCCCAACATCGAGGAAGTGGCCCTGTCCAACACTGGAGAGATTCCCTTCTAT 4457
Qy 4141 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4200
Db 4458 GGCAAAGCCATCCCCATTGAGGCCATCAAGGGGGGAAGGCATCTCATCTTCTGCCATTCC 4517
Qy 4201 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4260
Db 4518 AAGAAGAAGTGTGACGAGCTCGCCGCAAAGCTGACAGGCCTCGGACTCAATGCTGTAGCG 4577
Qy 4261 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4320
Db 4578 TATTACCGGGGTCTCGATGTGTCCGTCATACCGACTAGCGGAGACGTCGTTGTCGTGGCA 4637
Qy 4321 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4380
Db 4638 ACAGACGCTCTAATGACGGGTTTTACCGGCGACTTTGACTCAGTGATCGACTGCAACACA 4697
Qy 4381 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4440
Db 4698 TGTGTCACCCAGACAGTCGATTTCAGCTTGGATCCCACCTTCACCATTGAGACGACAACG 4757
Qy 4441 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4500
Db 4758 CTGCCCCAAGACGCGGTGTCGCGTGCGCAGCGGCGAGGTAGGACTGGCAGGGGCAGGAGT 4817
Qy 4501 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4560
Db 4818 GGCATCTACAGGTTTGTGACTCCAGGAGAACGGCCCTCAGGCATGTTCGACTCCTCGGTC 4877
Qy 4561 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4620
Db 4878 CTGTGTGAGTGCTATGACGCAGGCTGCGCTTGGTATGAGCTCACGCCCGCTGAGACCTCG 4937
Qy 4621 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 4680
Db 4938 GTTAGGTTGCGGGCTTACCTAAATACACCAGGGTTGCCCGTCTGCCAGGACCACCTAGAG 4997
Qy 4681 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 4740
Db 4998 TTCTGGGAGAGCGTCTTCACAGGCCTCACCCACATAGATGCCCACTTCTTGTCCCAGACC 5057
Qy 4741 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 4800
Db 5058 AAACAGGCAGGAGACAACCTCCCCTACCTGGTAGCATACCAAGCCACAGTGTGCGCCAGG 5117
Qy 4801 GCCAAAGCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 4860
Db 5118 GCTCAGGCTCCACCTCCATCGTGGGACCAAATGTGGAAGTGTCTCATACGGCTAAAGCCC 5177
Qy 4861 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 4920
Db 5178 ACACTGCATGGGCCAACGCCCCTGCTGTACAGGCTAGGAGCCGTTCAAAATGAGGTCACT 5237
Qy 4921 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 4980
Db 5238 CTCACACACCCCATAACCAAATACATCATGGCATGCATGTCGGCTGACCTGGAGGTCGTC 5297
Qy 4981 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5040
Db 5298 ACTAGCACCTGGGTGCTAGTAGGCGGAGTCCTTGCGGCTCTGGCCGCGTACTGCCTGACG 5357
Qy 5041 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5100
Db 5358 ACAGGCAGCGTGGTCATTGTGGGCAGGATCATCTTGTCCGGGAGGCCAGCTGTTATTCCC 5417
Qy 5101 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5160
Db 5418 GACAGGGAAGTCCTCTACCAGGAGTTCGATGAGATGGAAGAGTGTGCTTCACACCTCCCT 5477
Qy 5161 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5220
Db 5478 TACATCGAGCAAGGAATGCAGCTCGCCGAGCAATTCAAACAGAAGGCGCTCGGATTGCTG 5537
Qy 5221 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5280
Db 5538 CAAACAGCCACCAAGCAAGCGGAGGCTGCTGCTCCCGTGGTGGAGTCCAAGTGGCGAGCC 5597
Qy 5281 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5340
Db 5598 CTTGAGGTCTTCTGGGCGAAACACATGTGGAACTTCATCAGCGGGATACAGTACTTGGCA 5657
Qy 5341 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5400
Db 5658 GGCCTATCCACTCTGCCTGGAAACCCCGCGATAGCATCATTGATGGCTTTTACAGCCTCT 5717
Qy 5401 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5460
Db 5718 ATCACCAGCCCGCTCACCACCCAAAATACCCTCCTGTTTAACATCTTGGGGGGATGGGTG 5777

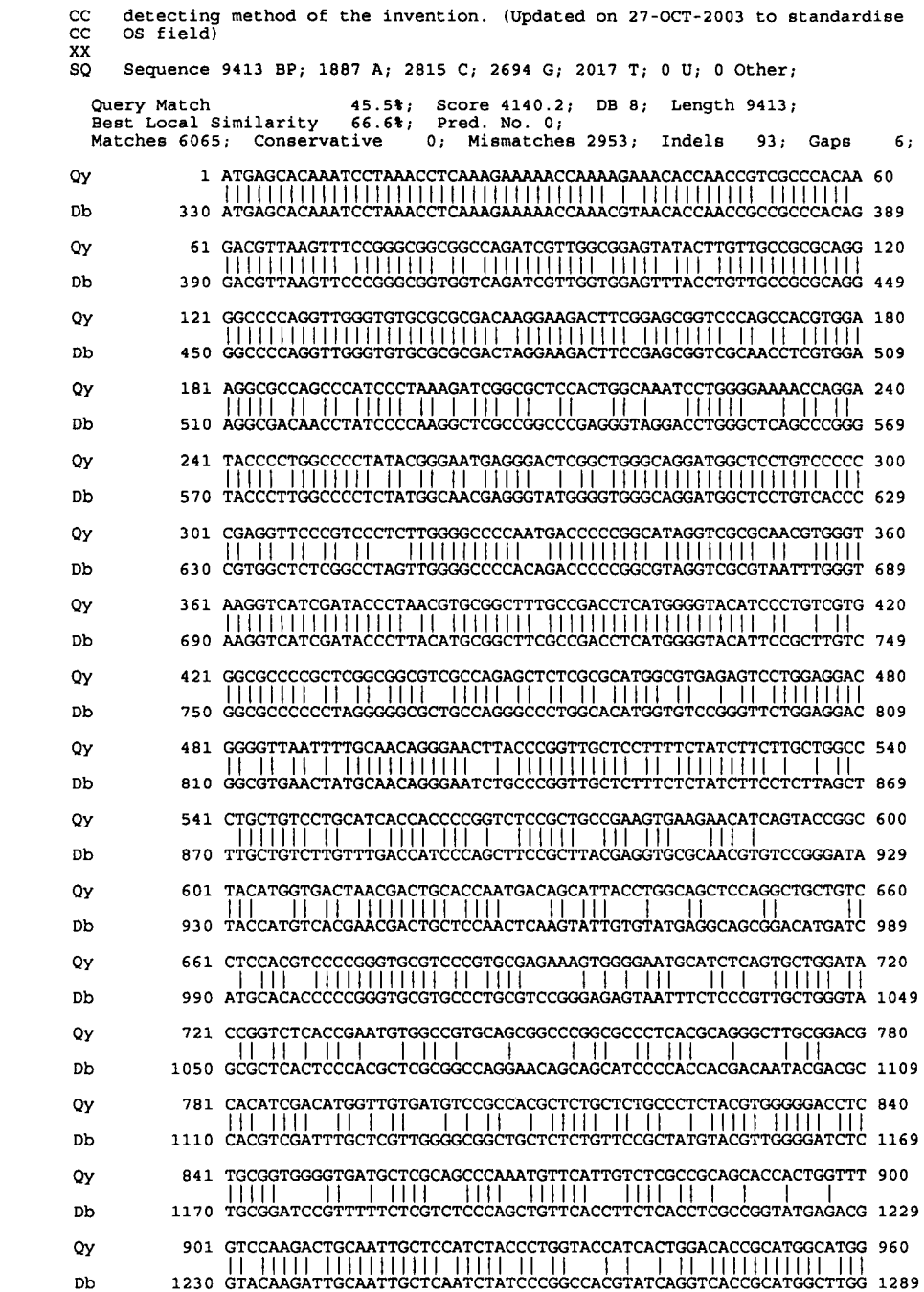

Qy 5461 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5520
Db 5778 GCTGCCCAACTCGCTCCCCCCAGCGCTGCTTCGGCTTTCGTGGGCGCCGGCATTGCCGGT 5837
Qy 5521 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5580
Db 5838 GCGGCCGTTGGCAGCATAGGTCTCGGGAAGGTACTTGTGGACATTCTGGCGGGCTATGGG 5897
Qy 5581 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5640
Db 5898 GCGGGGGTGGCTGGCGCACTCGTGGCCTTTAAGGTCATGAGCGGCGAGATGCCCTCCACT 5957
Qy 5641 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 5700
Db 5958 GAGGATCTGGTTAATTTACTCCCTGCCATCCTTTCTCCTGGCGCCCTGGTTGTCGGGGTC 6017
Qy 5701 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 5760
Db 6018 GTGTGCGCAGCAATACTGCGTCGGCACGTGGGCCCGGGAGAGGGGGCTGTGCAGTGGATG 6077
Qy 5761 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 5820
Db 6078 AACCGGCTGATAGCGTTCGCTTCGCGGGGTAACCACGTCTCCCCCACGCACTATGTGCCC 6137
Qy 5821 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 5880
Db 6138 GAGAGCGACGCCGCGGCGCGTGTTACTCAGATCCTCTCCAGCCTTACCATCACTCAGTTG 6197
Qy 5881 CTCAGAAGACTCCACAACTGGATTACTGAGGACTGCCCCATCCCATGCGGCGGCTCGTGG 5940
Db 6198 CTGAAGAGGCTTCATCAGTGGATTAATGAGGACTGCTCCACGCCTTGTTCCGGCTCGTGG 6257
Qy 5941 CTCCGCGATGTGTGGGACTGGGTTTGCACCATCCTAACAGACTTTAAAAATTGGCTGACC 6000
Db 6258 CTAAAGGATGTTTGGGACTGGATATGCACGGTGTTGAGTGACTTCAAGACTTGGCTCCAG 6317
Qy 6001 TCCAAATTATTCCCAAAGATGCCCGGCCTCCCCTTTGTCTCCTGTCAAAAGGGGTACAAG 6060
Db 6318 TCCAAGCTCCTGCCGCGGTTACCGGGACTCCCTTTCCTGTCATGCCAACGCGGGTACAAG 6377
Qy 6061 GGCGTGTGGGCCGGCACTGGCATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6120
Db 6378 GGAGTCTGGCGGGGGGATGGCATCATGCAAACCACCTGCCCATGTGGAGCACAGATCACC 6437
Qy 6121 GGCAATGTCCGCTTGGGCTCCATGAGAATCACGGGGCCTAAGACCTGCATGAATATCTGG 6180
Db 6438 GGACATGTCAAAAATGGCTCCATGAGGATTGTTGGGCCAAAAACCTGCAGCAACACGTGG 6497
Qy 6181 CAGGGGACCTTTCCTATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCGCCA 6240
Db 6498 CATGGAACATTCCCCATCAACGCATACACCACGGGCCCCTGCACGCCCTCCCCAGCGCCG 6557
Qy 6241 AACTTTAAGGTCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6300
Db 6558 AACTATTCCAGGGCGCTGTGGCGGGTGGCTGCTGAGGAGTACGTGGAGGTTACGCGGGTG 6617
Qy 6301 GGGTCATACCACTACATAACAGGACTCACCACTGATAACTTGAAAGTCCCCTGCCAACTA 6360
Db 6618 GGGGATTTCCACTACGTGACGGGCATGACCACTGACAACGTGAAATGCCCATGCCAGGTT 6677
Qy 6361 CCCTCTCCCGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCACA 6420
Db 6678 CCAGCCCCTGAATTTTTCACGGAGGTGGATGGAGTACGGTTGCACAGGTATGCTCCAGTG 6737
Qy 6421 CCGAAGCCGTTTTTCCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6480
Db 6738 TGCAAACCTCTCCTACGAGAGGAGGTCGTATTCCAGGTCGGGCTCAACCAGTACCTGGTC 6797
Qy 6481 GGGTCCCAGCTTCCTTGCGACCCTGAACCCGACACAGACGTATTGATGTCCATGCTAACA 6540
Db 6798 GGGTCACAGCTCCCATGTGAGCCCGAACCGGATGTGGCAGTGCTCACTTCCATGCTCACC 6857
Qy 6541 GATCCATCTCATATCACGGCGGAGACTGCAGCGCGGCGTTTAGCGCGGGGGTCACCCCCA 6600
Db 6858 GACCCCTCTCATATTACAGCAGAGACGGCCAAGCGTAGGCTGGCCAGGGGGTCTCCCCCC 6917
Qy 6601 TCCGAGGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 6660
Db 6918 TCCTTGGCCAGCTCTTCAGCTAGCCAGTTGTCTGCGCCTTCTTTGAAGGCGACATGTACT 6977
Qy 6661 ACCCACGGCAAAGCCTATGATGTGGACATGGTGGATGCTAACCT----------GTTC 6708
Db 6978 ACCCATCATGACTCCCCGGACGCTGACCTCATCGAGGCCAACCTCCTGTGGCGGCAGGAG 7037
Qy 6709 ATGGGGGGCGATGTGACTCGGATAGAGTCTGGGTCCAAAGTGGTCGTTCTGGACTCTCTC 6768
Db 7038 ATGGGCGGGAACATCACCCGTGTGGAGTCAGAAAATAAGGTGGTAATCCTGGACTCTTTC 7097
Qy 6769 GACCCAATGGTCGAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCAGAATACATGCTC 6828
Db 7098 GATCCGATTCGGGCGGTGGAGGATGAGAGGGAAATATCCGTCCCGGCGGAGATCCTGCGA 7157
Qy 6829 CCCAAGAAGAGGTTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCG 6888
Db 7158 AAACCCAGGAAGTTCCCCCCAGCGTTGCCCATATGGGCACGCCCGGATTACAACCCTCCA 7217
Qy 6889 CTTGTGGAATCGTGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTC 6948
Db 7218 CTGCTAGAGTCCTGGAAGGACCCGGACTACGTCCCCCCGGTGGTACACGGGTGCCCTTTG 7277
Qy 6949 CCTCCTCCTAGGAAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGT 7008
Db 7278 CCATCTACCAAGGCCCCCCCAATACCACCTCCACGGAGGAAGAGGACGGTTGTCCTGACA 7337
Qy 7009 GAGGACTCCATAGGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCC 7068
Db 7338 GAGTCCACCGTGTCTTCTGCCTTGGCGGAGCTCGCTACTAAGACCTTTGGCAGCTCCGGG 7397
Qy 7069 CCAAGCGGCGATTCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACG 7128
Db 7398 TCGTCGGCCGTTGACAGCGG---CACGGCGACTGGCCCTCCCGA---------TCAGGCC 7445
Qy 7129 CCTCCTGATGAGTTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCTCGAG 7188
Db 7446 TCCGACGACGGCGACAAAGGATCCGACGTTGAGTCGTACTCCTCCATGCCCCCCCTCGAG 7505
Qy 7189 GGGGAGCTTGGAGATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCAG 7248
Db 7506 GGAGAGCCAGGGGACCCCGACCT---------- 7528
Qy 7249 GGGGGGGTGGCAGCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGAC 7308
Db 7529 ----------CAGCGACGGGTCTTGGTCTACCGTGAGCGGGGAAGCTGGT 7568
Qy 7309 GACTCCGTCGTGTGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGT 7368
Db 7569 GAGGACGTCGTCTGCTGCTCAATGTCCTATACATGGACAGGTGCCTTGATCACGCCATGC 7628
Qy 7369 AGTCCCGAAGAGGAGAAGTTACCGATTAACCCCTTGAGCAACTCCCTGTTGCGATATCAC 7428
Db 7629 GCTGCGGAGGAGAGCAAGTTGCCCATCAATCCGTTGAGCAACTCTTTGCTGCGTCACCAC 7688
Qy 7429 AACAAGGTGTACTGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTT 7488
Db 7689 AGTATGGTCTACTCCACAACATCTCGCAGCGCAAGTCTGCGGCAGAAGAAGGTCACCTTT 7748
Qy 7489 GATAGGATGCAAGTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCG 7548
Db 7749 GACAGACTGCAAGTCCTGGACGACCACTACCGGGACGTGCTCAAGGAGATGAAGGCGAAG 7808
Qy 7549 GCCTCCAAGGTCACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCC 7608
Db 7809 GCGTCCACAGTTAAGGCTAGGCTTCTATCTATAGAGGAGGCCTGCAAACTGACGCCCCCA 7868
Qy 7609 CATTCTGCAAGATCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGG 7668

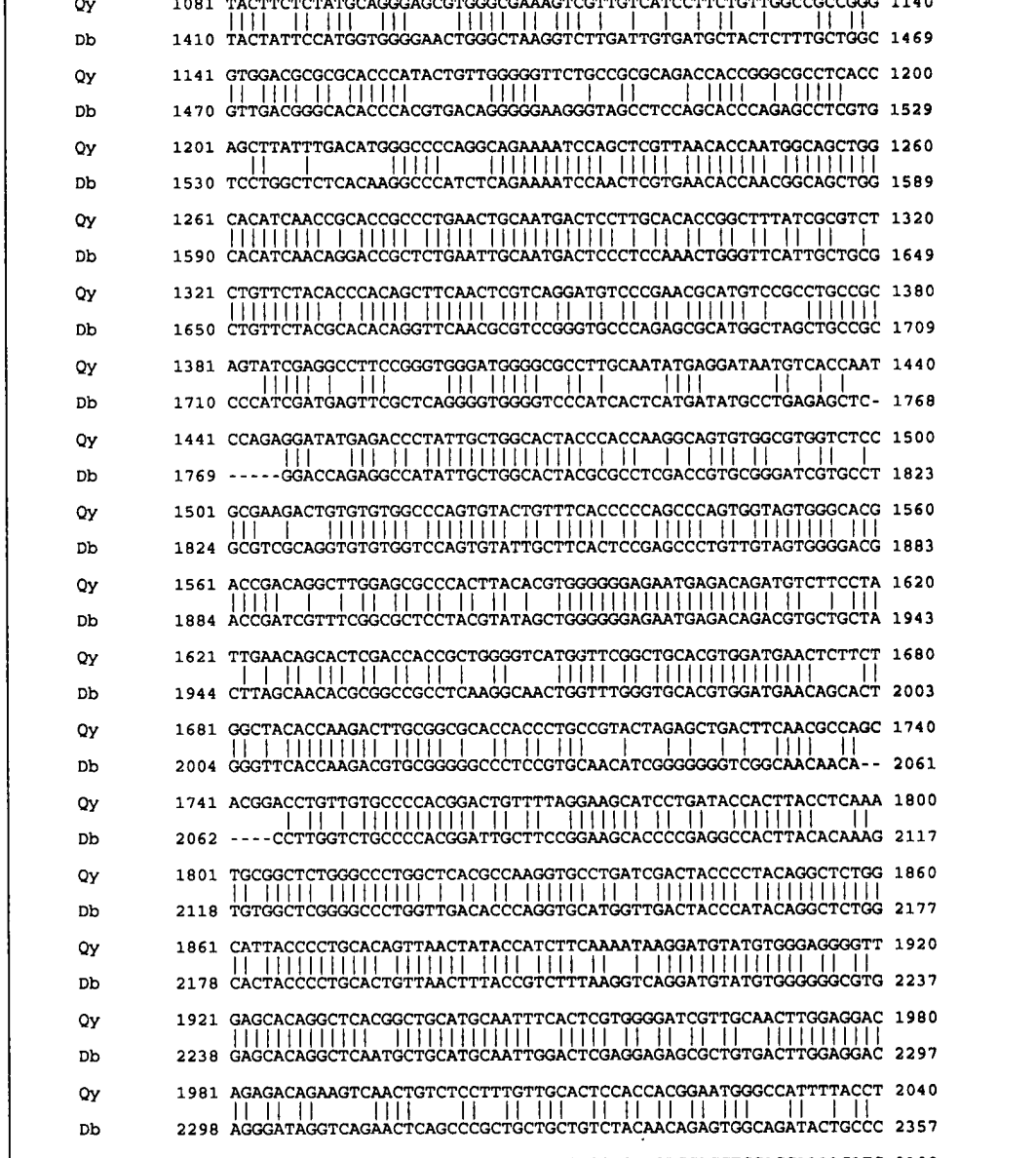


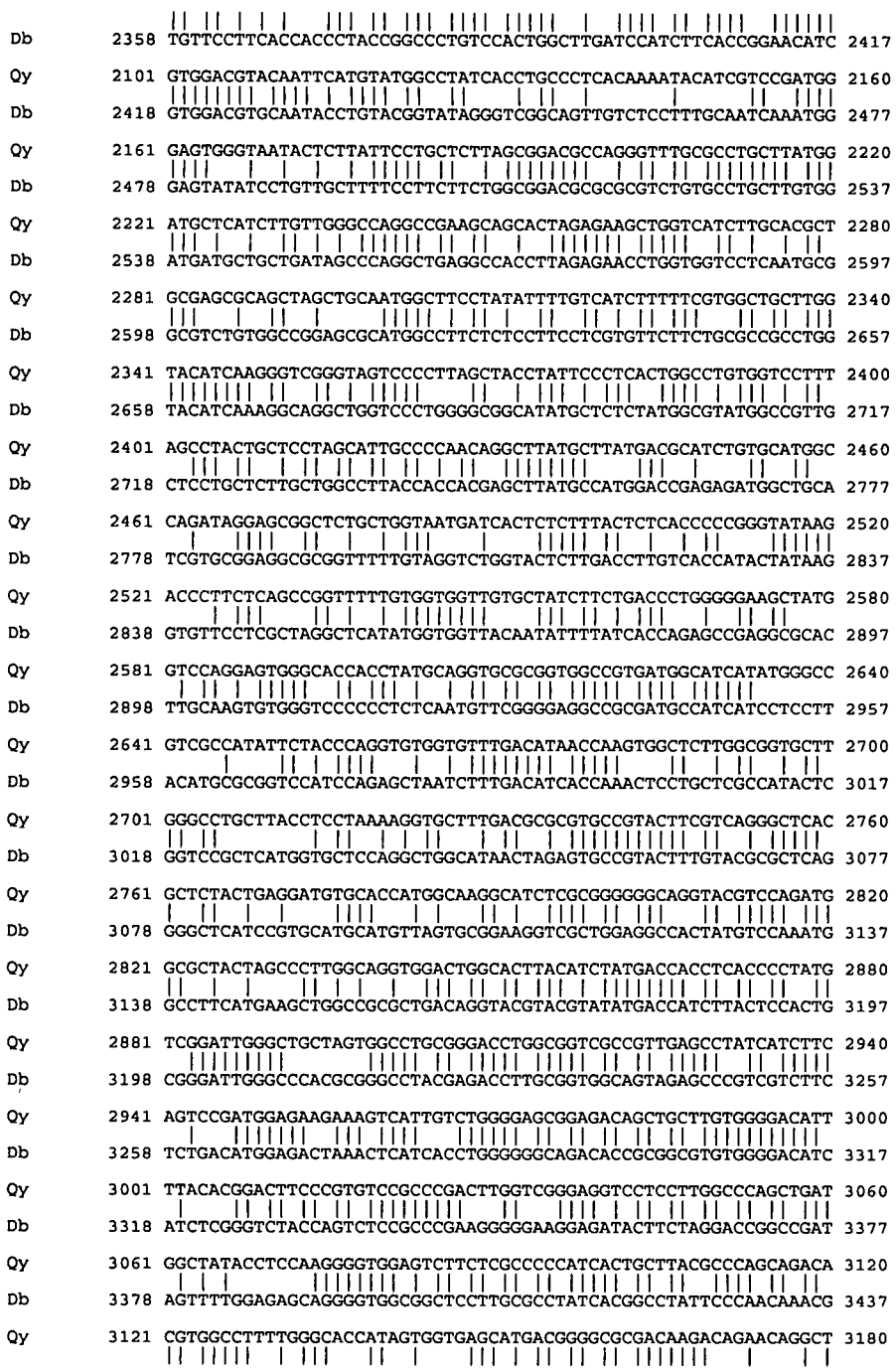
RESULT 15
AAL53723
ID AAL53723 standard; DNA; 9413 BP.
XX
AC AAL53723;
XX
DT 27-OCT-2003 (revised)
DT 07-FEB-2003 (first entry)
XX
DE Hepatitis C Virus DNA sequence SEQ ID No 17.
XX
KW Target RNA; target RNA:support-attached test compound; flow cytometry;
KW mass spectrometry; high-throughput screening; ds.
XX
OS Hepatitis C virus; Virus.
XX
PN WO200283837-A1.
XX
PD 24-OCT-2002.
XX
PF 11-APR-2002; 2002WO-US011758.
XX
PR 11-APR-2001; 2001US-0282966P.
XX
PA (PTCT-) PTC THERAPEUTICS INC.
XX
PI Almstead NG;
XX
DR WPI; 2003-075534/07.
XX
PT Identifying a test compound that binds to a target RNA molecule by
PT separating the detectably labeled target RNA:support-attached test
PT compound complex from uncomplexed target RNA molecules and test compounds
PT by flow cytometry.
XX
PS Example; Page 55-59; 131pp; English.
XX
CC The invention relates to a novel method for identifying a test compound
CC that binds to a target RNA molecule comprising separating the detectably
CC labeled target RNA:support-attached test compound complex from
CC uncomplexed target RNA molecules and test compounds. The separating
CC process is carried out by flow cytometry and determining a structure of
CC the type of test compound of the RNA:support-attached test compound
CC complex by mass spectrometry. The method is useful for high-throughput
CC screening of libraries of compounds to identify pharmaceutical leads.
CC This polynucleotide sequence represents a DNA sequence related to the

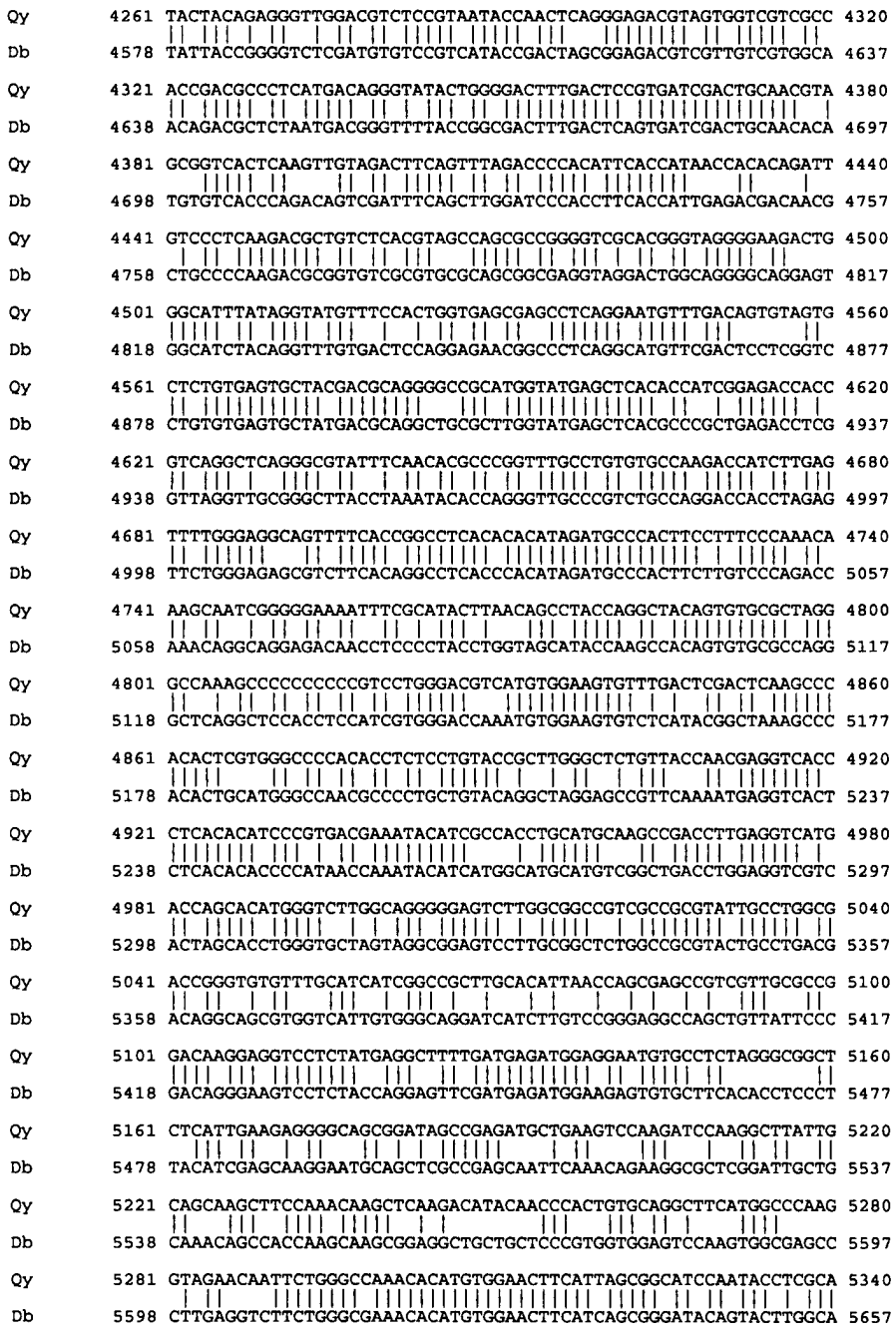
CC detecting method of the invention. (Updated on 27-OCT-2003 to standardise
CC OS field)
XX
SQ Sequence 9413 BP; 1887 A; 2815 C; 2694 G; 2017 T; 0 U; 0 Other;

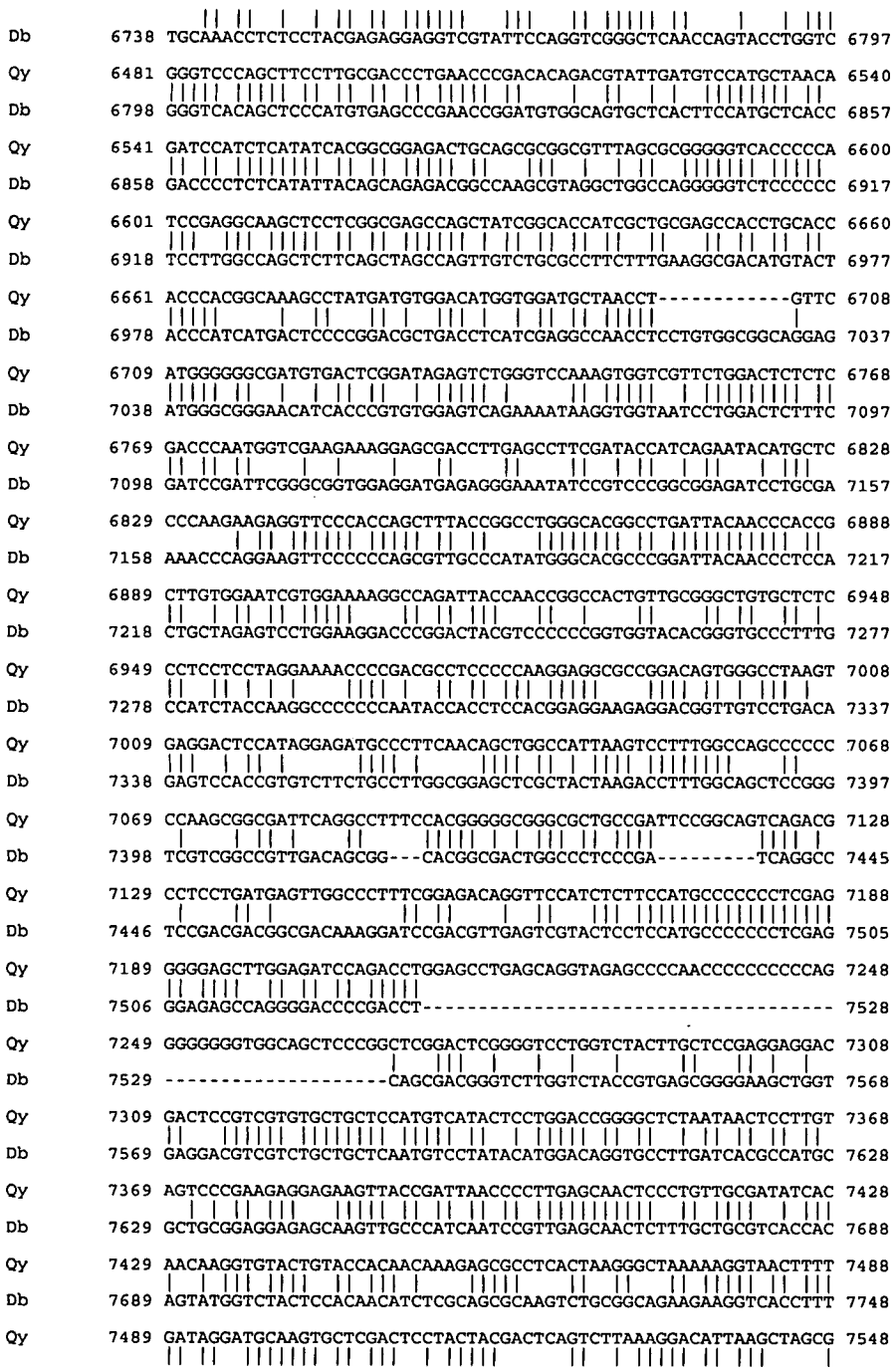
Query Match 45.5%; Score 4140.2; DB 8; Length 9413;
Best Local Similarity 66.6%; Pred. No. 0;
Matches 6065; Conservative 0; Mismatches 2953; Indels 93; Gaps 6;






Db 2358 TGTTCCTTCACCACCCTACCGGCCCTGTCCACTGGCTTGATCCATCTTCACCGGAACATC 2417
Qy 2101 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2160
Db 2418 GTGGACGTGCAATACCTGTACGGTATAGGGTCGGCAGTTGTCTCCTTTGCAATCAAATGG 2477
Qy 2161 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2220
Db 2478 GAGTATATCCTGTTGCTTTTCCTTCTTCTGGCGGACGCGCGCGTCTGTGCCTGCTTGTGG 2537
Qy 2221 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2280
Db 2538 ATGATGCTGCTGATAGCCCAGGCTGAGGCCACCTTAGAGAACCTGGTGGTCCTCAATGCG 2597
Qy 2281 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATATTTTGTCATCTTTTTCGTGGCTGCTTGG 2340
Db 2598 GCGTCTGTGGCCGGAGCGCATGGCCTTCTCTCCTTCCTCGTGTTCTTCTGCGCCGCCTGG 2657
Qy 2341 TACATCAAGGGTCGGGTAGTCCCCTTAGCTACCTATTCCCTCACTGGCCTGTGGTCCTTT 2400
Db 2658 TACATCAAAGGCAGGCTGGTCCCTGGGGCGGCATATGCTCTCTATGGCGTATGGCCGTTG 2717
Qy 2401 AGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTGCATGGC 2460
Db 2718 CTCCTGCTCTTGCTGGCCTTACCACCACGAGCTTATGCCATGGACCGAGAGATGGCTGCA 2777
Qy 2461 CAGATAGGAGCGGCTCTGCTGGTAATGATCACTCTCTTTACTCTCACCCCCGGGTATAAG 2520
Db 2778 TCGTGCGGAGGCGCGGTTTTTGTAGGTCTGGTACTCTTGACCTTGTCACCATACTATAAG 2837
Qy 2521 ACCCTTCTCAGCCGGTTTTTGTGGTGGTTGTGCTATCTTCTGACCCTGGGGGAAGCTATG 2580
Db 2838 GTGTTCCTCGCTAGGCTCATATGGTGGTTACAATATTTTATCACCAGAGCCGAGGCGCAC 2897
Qy 2581 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGCATCATATGGGCC 2640
Db 2898 TTGCAAGTGTGGGTCCCCCCTCTCAATGTTCGGGGAGGCCGCGATGCCATCATCCTCCTT 2957
Qy 2641 GTCGCCATATTCTACCCAGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 2700
Db 2958 ACATGCGCGGTCCATCCAGAGCTAATCTTTGACATCACCAAACTCCTGCTCGCCATACTC 3017
Qy 2701 GGGCCTGCTTACCTCCTAAAAGGTGCTTTGACGCGCGTGCCGTACTTCGTCAGGGCTCAC 2760
Db 3018 GGTCCGCTCATGGTGCTCCAGGCTGGCATAACTAGAGTGCCGTACTTTGTACGCGCTCAG 3077
Qy 2761 GCTCTACTGAGGATGTGCACCATGGCAAGGCATCTCGCGGGGGGCAGGTACGTCCAGATG 2820
Db 3078 GGGCTCATCCGTGCATGCATGTTAGTGCGGAAGGTCGCTGGAGGCCACTATGTCCAAATG 3137
Qy 2821 GCGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 2880
Db 3138 GCCTTCATGAAGCTGGCCGCGCTGACAGGTACGTACGTATATGACCATCTTACTCCACTG 3197
Qy 2881 TCGGATTGGGCTGCTAGTGGCCTGCGGGACCTGGCGGTCGCCGTTGAGCCTATCATCTTC 2940
Db 3198 CGGGATTGGGCCCACGCGGGCCTACGAGACCTTGCGGTGGCAGTAGAGCCCGTCGTCTTC 3257
Qy 2941 AGTCCGATGGAGAAGAAAGTCATTGTCTGGGGAGCGGAGACAGCTGCTTGTGGGGACATT 3000
Db 3258 TCTGACATGGAGACTAAACTCATCACCTGGGGGGCAGACACCGCGGCGTGTGGGGACATC 3317
Qy 3001 TTACACGGACTTCCCGTGTCCGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3060
Db 3318 ATCTCGGGTCTACCAGTCTCCGCCCGAAGGGGGAAGGAGATACTTCTAGGACCGGCCGAT 3377
Qy 3061 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3120
Db 3378 AGTTTTGGAGAGCAGGGGTGGCGGCTCCTTGCGCCTATCACGGCCTATTCCCAACAAACG 3437
Qy 3121 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCT 3180
Db 3438 CGGGGCCTGCTTGGCTGTATCATCACTAGCCTCACAGGTCGGGACAAGAACCAGGTCGAT 3497
Qy 3181 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3240
Db 3498 GGGGAGGTTCAGGTGCTCTCCACCGCAACGCAATCTTTCCTGGCGACCTGCGTCAATGGC 3557
Qy 3241 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3300
Db 3558 GTGTGTTGGACCGTCTACCATGGTGCCGGCTCGAAGACCCTGGCCGGCCCGAAGGGTCCA 3617
Qy 3301 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGG 3360
Db 3618 ATCACCCAAATGTACACCAATGTAGACCAGGACCTCGTCGGCTGGCCGGCGCCCCCCGGG 3677
Qy 3361 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3420
Db 3678 GCGCGCTCCATGACACCGTGCACCTGCGGCAGCTCGGACCTTTACTTGGTCACGAGGCAT 3737
Qy 3421 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3480
Db 3738 GCTGATGTCGTTCCGGTGCGCCGGCGGGGCGACAGCAGGGGGAGCCTGCTTTCCCCCAGG 3797
Qy 3481 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3540
Db 3798 CCCATCTCCTACCTGAAGGGCTCCTCGGGTGGACCACTGCTTTGCCCTTCGGGGCACGTT 3857
Qy 3541 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3600
Db 3858 GTAGGCATCTTCCGGGCTGCTGTGTGCACCCGGGGGGTTGCGAAGGCGGTGGACTTCATA 3917
Qy 3601 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 3660
Db 3918 CCCGTTGAGTCTATGGAAACTACCATGCGGTCTCCGGTCTTCACAGACAACTCATCCCCT 3977
Qy 3661 CCTGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 3720
Db 3978 CCGGCCGTACCGCAAACATTCCAAGTGGCACATTTACACGCTCCCACTGGCAGCGGCAAG 4037
Qy 3721 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 3780
Db 4038 AGCACCAAAGTGCCGGCTGCATATGCAGCCCAAGGGTACAAGGTGCTCGTCCTAAACCCG 4097
Qy 3781 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 3840
Db 4098 TCCGTTGCCGCCACATTGGGCTTTGGAGCGTATATGTCCAAGGCACATGGCATCGAGCCT 4157
Qy 3841 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 3900
Db 4158 AACATCAGAACTGGGGTAAGGACCATCACCACGGGCGGCCCCATCACGTACTCCACCTAT 4217
Qy 3901 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 3960
Db 4218 TGCAAGTTCCTTGCCGACGGTGGATGCTCCGGGGGCGCCTATGACATCATAATATGTGAT 4277
Qy 3961 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4020
Db 4278 GAATGCCACTCAACTGACTCGACTACCATCTTGGGCATCGGCACAGTCCTGGATCAGGCA 4337
Qy 4021 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4080
Db 4338 GAGACGGCTGGAGCGCGGCTCGTCGTGCTCGCCACCGCCACGCCTCCGGGATCGATCACC 4397
Qy 4081 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4140
Db 4398 GTGCCACACCCCAACATCGAGGAAGTGGCCCTGTCCAACACTGGAGAGATTCCCTTCTAT 4457
Qy 4141 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4200
Db 4458 GGCAAAGCCATCCCCATTGAGGCCATCAAGGGGGGAAGGCATCTCATCTTCTGCCATTCC 4517
Qy 4201 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4260
Db 4518 AAGAAGAAGTGTGACGAGCTCGCCGCAAAGCTGACAGGCCTCGGACTCAATGCTGTAGCG 4577


Qy 4261 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4320
Db 4578 TATTACCGGGGTCTCGATGTGTCCGTCATACCGACTAGCGGAGACGTCGTTGTCGTGGCA 4637
Qy 4321 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4380
Db 4638 ACAGACGCTCTAATGACGGGTTTTACCGGCGACTTTGACTCAGTGATCGACTGCAACACA 4697
Qy 4381 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4440
Db 4698 TGTGTCACCCAGACAGTCGATTTCAGCTTGGATCCCACCTTCACCATTGAGACGACAACG 4757
Qy 4441 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4500
Db 4758 CTGCCCCAAGACGCGGTGTCGCGTGCGCAGCGGCGAGGTAGGACTGGCAGGGGCAGGAGT 4817
Qy 4501 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4560
Db 4818 GGCATCTACAGGTTTGTGACTCCAGGAGAACGGCCCTCAGGCATGTTCGACTCCTCGGTC 4877
Qy 4561 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4620
Db 4878 CTGTGTGAGTGCTATGACGCAGGCTGCGCTTGGTATGAGCTCACGCCCGCTGAGACCTCG 4937
Qy 4621 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 4680
Db 4938 GTTAGGTTGCGGGCTTACCTAAATACACCAGGGTTGCCCGTCTGCCAGGACCACCTAGAG 4997
Qy 4681 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 4740
Db 4998 TTCTGGGAGAGCGTCTTCACAGGCCTCACCCACATAGATGCCCACTTCTTGTCCCAGACC 5057
Qy 4741 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 4800
Db 5058 AAACAGGCAGGAGACAACCTCCCCTACCTGGTAGCATACCAAGCCACAGTGTGCGCCAGG 5117
Qy 4801 GCCAAAGCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 4860
Db 5118 GCTCAGGCTCCACCTCCATCGTGGGACCAAATGTGGAAGTGTCTCATACGGCTAAAGCCC 5177
Qy 4861 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 4920
Db 5178 ACACTGCATGGGCCAACGCCCCTGCTGTACAGGCTAGGAGCCGTTCAAAATGAGGTCACT 5237
Qy 4921 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 4980
Db 5238 CTCACACACCCCATAACCAAATACATCATGGCATGCATGTCGGCTGACCTGGAGGTCGTC 5297
Qy 4981 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5040
Db 5298 ACTAGCACCTGGGTGCTAGTAGGCGGAGTCCTTGCGGCTCTGGCCGCGTACTGCCTGACG 5357
Qy 5041 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5100
Db 5358 ACAGGCAGCGTGGTCATTGTGGGCAGGATCATCTTGTCCGGGAGGCCAGCTGTTATTCCC 5417
Qy 5101 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5160
Db 5418 GACAGGGAAGTCCTCTACCAGGAGTTCGATGAGATGGAAGAGTGTGCTTCACACCTCCCT 5477
Qy 5161 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5220
Db 5478 TACATCGAGCAAGGAATGCAGCTCGCCGAGCAATTCAAACAGAAGGCGCTCGGATTGCTG 5537
Qy 5221 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5280
Db 5538 CAAACAGCCACCAAGCAAGCGGAGGCTGCTGCTCCCGTGGTGGAGTCCAAGTGGCGAGCC 5597
Qy 5281 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5340
Db 5598 CTTGAGGTCTTCTGGGCGAAACACATGTGGAACTTCATCAGCGGGATACAGTACTTGGCA 5657
Qy 5341 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5400
Db 5658 GGCCTATCCACTCTGCCTGGAAACCCCGCGATAGCATCATTGATGGCTTTTACAGCCTCT 5717
Qy 5401 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5460
Db 5718 ATCACCAGCCCGCTCACCACCCAAAATACCCTCCTGTTTAACATCTTGGGGGGATGGGTG 5777
Qy 5461 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5520
Db 5778 GCTGCCCAACTCGCTCCCCCCAGCGCTGCTTCGGCTTTCGTGGGCGCCGGCATTGCCGGT 5837
Qy 5521 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5580
Db 5838 GCGGCCGTTGGCAGCATAGGTCTCGGGAAGGTACTTGTGGACATTCTGGCGGGCTATGGG 5897
Qy 5581 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5640
Db 5898 GCGGGGGTGGCTGGCGCACTCGTGGCCTTTAAGGTCATGAGCGGCGAGATGCCCTCCACT 5957
Qy 5641 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 5700
Db 5958 GAGGATCTGGTTAATTTACTCCCTGCCATCCTTTCTCCTGGCGCCCTGGTTGTCGGGGTC 6017
Qy 5701 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 5760
Db 6018 GTGTGCGCAGCAATACTGCGTCGGCACGTGGGCCCGGGAGAGGGGGCTGTGCAGTGGATG 6077
Qy 5761 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 5820
Db 6078 AACCGGCTGATAGCGTTCGCTTCGCGGGGTAACCACGTCTCCCCCACGCACTATGTGCCC 6137
Qy 5821 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 5880
Db 6138 GAGAGCGACGCCGCGGCGCGTGTTACTCAGATCCTCTCCAGCCTTACCATCACTCAGTTG 6197
Qy 5881 CTCAGAAGACTCCACAACTGGATTACTGAGGACTGCCCCATCCCATGCGGCGGCTCGTGG 5940
Db 6198 CTGAAGAGGCTTCATCAGTGGATTAATGAGGACTGCTCCACGCCTTGTTCCGGCTCGTGG 6257
Qy 5941 CTCCGCGATGTGTGGGACTGGGTTTGCACCATCCTAACAGACTTTAAAAATTGGCTGACC 6000
Db 6258 CTAAAGGATGTTTGGGACTGGATATGCACGGTGTTGAGTGACTTCAAGACTTGGCTCCAG 6317
Qy 6001 TCCAAATTATTCCCAAAGATGCCCGGCCTCCCCTTTGTCTCCTGTCAAAAGGGGTACAAG 6060
Db 6318 TCCAAGCTCCTGCCGCGGTTACCGGGACTCCCTTTCCTGTCATGCCAACGCGGGTACAAG 6377
Qy 6061 GGCGTGTGGGCCGGCACTGGCATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6120
Db 6378 GGAGTCTGGCGGGGGGATGGCATCATGCAAACCACCTGCCCATGTGGAGCACAGATCACC 6437
Qy 6121 GGCAATGTCCGCTTGGGCTCCATGAGAATCACGGGGCCTAAGACCTGCATGAATATCTGG 6180
Db 6438 GGACATGTCAAAAATGGCTCCATGAGGATTGTTGGGCCAAAAACCTGCAGCAACACGTGG 6497
Qy 6181 CAGGGGACCTTTCCTATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCGCCA 6240
Db 6498 CATGGAACATTCCCCATCAACGCATACACCACGGGCCCCTGCACGCCCTCCCCAGCGCCG 6557
Qy 6241 AACTTTAAGGTCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6300
Db 6558 AACTATTCCAGGGCGCTGTGGCGGGTGGCTGCTGAGGAGTACGTGGAGGTTACGCGGGTG 6617
Qy 6301 GGGTCATACCACTACATAACAGGACTCACCACTGATAACTTGAAAGTCCCCTGCCAACTA 6360
Db 6618 GGGGATTTCCACTACGTGACGGGCATGACCACTGACAACGTGAAATGCCCATGCCAGGTT 6677
Qy 6361 CCCTCTCCCGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCACA 6420
Db 6678 CCAGCCCCTGAATTTTTCACGGAGGTGGATGGAGTACGGTTGCACAGGTATGCTCCAGTG 6737
Qy 6421 CCGAAGCCGTTTTTCCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6480


Db 6738 TGCAAACCTCTCCTACGAGAGGAGGTCGTATTCCAGGTCGGGCTCAACCAGTACCTGGTC 6797
Qy 6481 GGGTCCCAGCTTCCTTGCGACCCTGAACCCGACACAGACGTATTGATGTCCATGCTAACA 6540
Db 6798 GGGTCACAGCTCCCATGTGAGCCCGAACCGGATGTGGCAGTGCTCACTTCCATGCTCACC 6857
Qy 6541 GATCCATCTCATATCACGGCGGAGACTGCAGCGCGGCGTTTAGCGCGGGGGTCACCCCCA 6600
Db 6858 GACCCCTCTCATATTACAGCAGAGACGGCCAAGCGTAGGCTGGCCAGGGGGTCTCCCCCC 6917
Qy 6601 TCCGAGGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 6660
Db 6918 TCCTTGGCCAGCTCTTCAGCTAGCCAGTTGTCTGCGCCTTCTTTGAAGGCGACATGTACT 6977
Qy 6661 ACCCACGGCAAAGCCTATGATGTGGACATGGTGGATGCTAACCT----------GTTC 6708
Db 6978 ACCCATCATGACTCCCCGGACGCTGACCTCATCGAGGCCAACCTCCTGTGGCGGCAGGAG 7037
Qy 6709 ATGGGGGGCGATGTGACTCGGATAGAGTCTGGGTCCAAAGTGGTCGTTCTGGACTCTCTC 6768
Db 7038 ATGGGCGGGAACATCACCCGTGTGGAGTCAGAAAATAAGGTGGTAATCCTGGACTCTTTC 7097
Qy 6769 GACCCAATGGTCGAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCAGAATACATGCTC 6828
Db 7098 GATCCGATTCGGGCGGTGGAGGATGAGAGGGAAATATCCGTCCCGGCGGAGATCCTGCGA 7157
Qy 6829 CCCAAGAAGAGGTTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCG 6888
Db 7158 AAACCCAGGAAGTTCCCCCCAGCGTTGCCCATATGGGCACGCCCGGATTACAACCCTCCA 7217
Qy 6889 CTTGTGGAATCGTGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTC 6948
Db 7218 CTGCTAGAGTCCTGGAAGGACCCGGACTACGTCCCCCCGGTGGTACACGGGTGCCCTTTG 7277
Qy 6949 CCTCCTCCTAGGAAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGT 7008
Db 7278 CCATCTACCAAGGCCCCCCCAATACCACCTCCACGGAGGAAGAGGACGGTTGTCCTGACA 7337
Qy 7009 GAGGACTCCATAGGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCC 7068
Db 7338 GAGTCCACCGTGTCTTCTGCCTTGGCGGAGCTCGCTACTAAGACCTTTGGCAGCTCCGGG 7397
Qy 7069 CCAAGCGGCGATTCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACG 7128
Db 7398 TCGTCGGCCGTTGACAGCGG---CACGGCGACTGGCCCTCCCGA---------TCAGGCC 7445
Qy 7129 CCTCCTGATGAGTTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCTCGAG 7188
Db 7446 TCCGACGACGGCGACAAAGGATCCGACGTTGAGTCGTACTCCTCCATGCCCCCCCTCGAG 7505
Qy 7189 GGGGAGCTTGGAGATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCAG 7248
Db 7506 GGAGAGCCAGGGGACCCCGACCT---------- 7528
Qy 7249 GGGGGGGTGGCAGCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGAC 7308
Db 7529 ----------CAGCGACGGGTCTTGGTCTACCGTGAGCGGGGAAGCTGGT 7568
Qy 7309 GACTCCGTCGTGTGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGT 7368
Db 7569 GAGGACGTCGTCTGCTGCTCAATGTCCTATACATGGACAGGTGCCTTGATCACGCCATGC 7628
Qy 7369 AGTCCCGAAGAGGAGAAGTTACCGATTAACCCCTTGAGCAACTCCCTGTTGCGATATCAC 7428
Db 7629 GCTGCGGAGGAGAGCAAGTTGCCCATCAATCCGTTGAGCAACTCTTTGCTGCGTCACCAC 7688
Qy 7429 AACAAGGTGTACTGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTT 7488
Db 7689 AGTATGGTCTACTCCACAACATCTCGCAGCGCAAGTCTGCGGCAGAAGAAGGTCACCTTT 7748
Qy 7489 GATAGGATGCAAGTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCG 7548
Db 7749 GACAGACTGCAAGTCCTGGACGACCACTACCGGGACGTGCTCAAGGAGATGAAGGCGAAG 7808
Qy 7549 GCCTCCAAGGTCACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCC 7608
Db 7809 GCGTCCACAGTTAAGGCTAGGCTTCTATCTATAGAGGAGGCCTGCAAACTGAAGCCCCCA 7868
Qy 7609 CATTCTGCAAGATCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGG 7668
Db 7869 CATTCGGCCAAATCCAAATTTGGCTACGGGGCGAAGGACGTCCGGAGCCTATCCAGCAGG 7928
Qy 7669 GCCGTTAACCACATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATT 7728
Db 7929 GCCGTCAACCACATCCGCTCCGTGTGGGAGGACTTGCTGGAAGACACTGAAACACCAATT 7988
Qy 7729 CCCACAACCATTATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAG 7788
Db 7989 GATACCACCATCATGGCAAAAAATGAGGTTTTCTGCGTCCAACCAGAGAAAGGAGGCCGC 8048
Qy 7789 AAAGCAGCTCGCCTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCC 7848
Db 8049 AAGCCAGCTCGCCTTATCGTATTCCCAGACCTGGGGGTACGTGTATGCGAGAAGATGGCC 8108
Qy 7849 CTTTATGACATTACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAG 7908
Db 8109 CTTTACGACGTGGTCTCCACCCTTCCTCAGGCCGTGATGGGCCCCTCATACGGATTCCAG 8168
Qy 7909 TATTCCCCCGCTCAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCT 7968
Db 8169 TACTCTCCTGGGCAGCGGGTCGAGTTCCTGGTGAATACCTGGAAATCAAAGAAATGCCCT 8228
Qy 7969 ATGGGTTTTTCGTATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGG 8028
Db 8229 ATGGGCTTCTCATATGACACCCGCTGCTTTGACTCAACGGTCACTGAGAATGACATCCGT 8288
Qy 8029 ACTGAGGAGTCCATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACAC 8088
Db 8289 ACTGAGGAATCAATTTACCAATGTTGTGACTTGGCCCCCGAAGCCAGGCAGGCCATAAGG 8348
Qy 8089 TCGCTAACTGAGAGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGC 8148
Db 8349 TCGCTCACAGAGCGGCTTTATGTCGGGGGTCCCCTGACTAATTCGAAGGGGCAGAACTGC 8408
Qy 8149 GGGTACAGGCGTTGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACA 8208
Db 8409 GGTTATCGCCGGTGCCGCGCAAGTGGCGTGCTGACGACTAGCTGCGGCAACACCCTCACA 8468
Qy 8209 TGCTACGTGAAAGCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTG 8268
Db 8469 TGTTACTTGAAGGCCACTGCGGCCTGTCGAGCTGCAAAGCTCCAGGACTGCACGATGCTC 8528
Qy 8269 GTATGCGGCGATGACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGG 8328
Db 8529 GTGAACGGAGACGACCTTGTCGTTATCTGTGAGAGTGCGGGAACCCAGGAGGATGCGGCG 8588
Qy 8329 AACCTGAGAGCCTTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCC 8388
Db 8589 GCCCTACGAGCCTTCACGGAGGCTATGACTAGGTATTCCGCCCCCCCCGGGGACCCGCCC 8648
Qy 8389 AGACCGGAGTATGATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTG 8448
Db 8649 CAACCAGAATACGACTTGGAGCTGATAACGTCATGCTCCTCCAATGTGTCGGTCGCGCAC 8708
Qy 8449 GGCCCACAAGGCCGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGG 8508
Db 8709 GATGCATCCGGCAAAAGGGTGTACTACCTCACCCGTGACCCCACCACCCCCCTCGCACGG 8768
Qy 8509 GCTGCCTGGGAAACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAG 8568
Db 8769 GCTGCGTGGGAGACAGTTAGACACACTCCAGTCAACTCCTGGCTAGGCAATATCATCATG 8828
Qy 8569 TACGCCCCGACCATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATG 8628
Db 8829 TATGCGCCCACCCTATGGGCGAGGATGATTCTGATGACTCATTTCTTCTCTATCCTTCTA 8888

```
Qy       8629 GCTCAAGACACGCTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTG 8688
              |||||  ||    || ||  |   |||  | | | |||| ||||| ||    ||||||
Db       8889 GCTCAGGAGCAACTTGAAAAAGCCCTGGATTGTCAGATCTACGGGGCCTGTTACTCCATT 8948

Qy       8689 AGTCCCTTGGACCTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTG 8748
                ||  | ||||| ||    || |||||| |  | ||||| |||  ||| ||||| ||
Db       8949 GAGCCACTTGACCTACCTCAGATCATTGAACGACTCCATGGTCTTAGCGCATTTTCACTC 9008

Qy       8749 CACACATACACTCCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCG 8808
              ||||  |||  ||||    ||  | |   |||||||||||  |||||| ||||||||||
Db       9009 CACAGTTACTCTCCAGGTGAGATCAATAGGGTGGCTTCATGCCTCAGGAAACTTGGGGTA 9068

Qy       8809 CCACCCCTCAGAGCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGT 8868
              || ||  |  |||  ||||     |||||  |    ||| | ||    ||  | ||||
Db       9069 CCGCCTTTGCGAGTCTGGAGACATCGGGCCAGAAGTGTCCGCGCTAAGCTACTGTCCCAG 9128

Qy       8869 GGGGGGAGAGCGGCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTC 8928
              |||||||| || |||  ||||||    || |||||||| ||||| || ||||||||||
Db       9129 GGGGGGAGGGCTGCCACTTGCGGCAAGTACCTCTTCAACTGGGCAGTAAAGACCAAGCTT 9188

Qy       8929 AAACTCACTCCATTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGC 8988
              ||||||||||| | ||||  |||  ||  || || || ||| ||||||||   |  ||
Db       9189 AAACTCACTCCAATCCCGGCTGCGTCCCAGCTAGACTTGTCCGGCTGGTTCGTTGCTGGT 9248

Qy       8989 GCCGGCGGGGGCGACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTT 9048
                |  |||||| ||||| |||||||||| |||| ||||||||||||||| |  ||  | ||
Db       9249 TACAACGGGGGAGACATATATCACAGCCTGTCTCGTGCCCGACCCCGTTGGTTCATGTTG 9308

Qy       9049 GGCCTACTCCTACTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCGG 9099
              |||||||||||||| |||||||||||| | | ||| ||||||   |||
Db       9309 TGCCTACTCCTACTTTCTGTAGGGGTAGGCATCTACCTGCTCCCCAACCGG 9359
```

Search completed: October 31, 2005, 16:23:32
Job time : 3057 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 31, 2005, 15:24:05 ; Search time 17316 Seconds
(without alignments)
20001.547 Million cell updates/sec

Title: US-09-980-559-1_COPY_341_9439
Perfect score: 9099
Sequence: 1 atgagcacaaatcctaaacc.....ttttcctactccccgctcgg 9099

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_htc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	274.6	3.0	488	1	AV755731	AV755731
c	2	235.8	2.6	492	1	AV758366	AV758366
	3	57.6	0.6	179	7	R28798	R28798 F0-266D 22
	4	52	0.6	522	7	CN373175	CN373175 170005313
	5	52	0.6	925	9	CNS0091P	AL053013 Drosophil
c	6	51.8	0.6	925	9	CNS0091P	AL053013 Drosophil
c	7	51.4	0.6	1203	9	CNS015Y4	AL106054 Drosophil
	8	50.6	0.6	4762	3	HSM804270	AL832959 Homo sapi
	9	50	0.5	3661	9	AY411015	AY411015 Homo sapi
c	10	49.2	0.5	844	9	CNS0052P	AL056652 Drosophil
	11	48.8	0.5	1082	5	BQ940004	BQ940004 AGENCOURT
c	12	47.2	0.5	393	8	BH622015	BH622015 1007116B0
	13	46	0.5	1009	9	CNS010EW	AL098882 Drosophil
c	14	45.8	0.5	755	7	CK307855	CK307855 SB02044B2
	15	45.6	0.5	562	6	CA733786	CA733786 wlp1c.pk0
	16	44.4	0.5	515	6	CA015275	CA015275 HT13N01r
	17	44.4	0.5	1201	9	CNS016BR	AL106545 Drosophil
	18	44.2	0.5	910	9	CNS006ON	AL065629 Drosophil
	19	44	0.5	486	6	CD938672	CD938672 OV.110K07
	20	44	0.5	536	5	BQ801777	BQ801777 WHE2818_D
	21	43.8	0.5	894	5	BX425557	BX425557 BX425557
	22	43.6	0.5	582	6	CD937954	CD937954 OV.108J14
	23	43.4	0.5	1101	9	CNS00LXJ	AL078875 Drosophil
	24	43.2	0.5	472	2	BF652214	BF652214 275714 MA
c	25	43.2	0.5	555	9	CG566920	CG566920 OST192889
	26	43	0.5	478	5	BQ460624	BQ460624 HA06G17r
	27	43	0.5	595	2	AW666752	AW666752 GA__Ea000
	28	43	0.5	650	6	CA070239	CA070239 SCSGAD100
	29	43	0.5	718	8	AZ211102	AZ211102 SP_0159_B
c	30	43	0.5	839	9	CNS004NB	AL054280 Drosophil
	31	42.6	0.5	385	7	CO954417	CO954417 UMC-pnata
	32	42.6	0.5	391	2	AW344508	AW344508 24363 MAR
	33	42.6	0.5	493	2	BF079917	BF079917 230626 MA
	34	42.6	0.5	506	4	BG895775	BG895775 359402 MA
	35	42.6	0.5	542	4	BG609855	BG609855 324097 MA
	36	42.6	0.5	549	2	AW784808	AW784808 115078 MA
	37	42.6	0.5	560	4	BG895608	BG895608 359140 MA
	38	42.6	0.5	581	4	BG733252	BG733252 347020 MA
	39	42.6	0.5	583	4	BI346523	BI346523 375709 MA
	40	42.6	0.5	748	2	BE558337	BE558337 HV_CEb001
c	41	42.4	0.5	442	1	AF367693	AF367693 AF367693
	42	42.4	0.5	562	6	CA671329	CA671329 wlsu2.pk0
	43	42.4	0.5	589	6	CA635359	CA635359 wleln.pk0
c	44	42.4	0.5	787	2	BE867644	BE867644 601443160
	45	42.4	0.5	907	5	BU859486	BU859486 AGENCOURT

ALIGNMENTS

RESULT 1
AV755731/c
LOCUS AV755731 488 bp mRNA linear EST 19-OCT-2000
DEFINITION AV755731 BM Homo sapiens cDNA clone BMFAKB03 5', mRNA sequence.
ACCESSION AV755731
VERSION AV755731.1 GI:10913579
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 488)
AUTHORS Gu,J., Zhao,M., Huang,Q., Xu,X., Li,Y., Peng,Y., Song,H., Xiao,H., Gu,Y., Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Yang,Y., Gao,G., Wang,Z., Zhang,Q., Chen,S., Han,Z. and Chen,Z.
TITLE Homo sapiens cDNA BM clones
JOURNAL Unpublished (2000)
COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES Location/Qualifiers
source 1. .488
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="BMFAKB03"
/tissue_type="Bone marrow"
/cell_type="CD34+ hematopoietic stem/progenitor cell"
/lab_host="BM25.8"
/clone_lib="BM"
/note="Vector: pTriplEx2; Site_1: sfiIA; Site_2: sfiIB"
ORIGIN

Query Match 3.0%; Score 274.6; DB 1; Length 488;
Best Local Similarity 78.4%; Pred. No. 2.5e-64;
Matches 366; Conservative 0; Mismatches 96; Indels 5; Gaps 3;

```
Qy        796 GTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTCTGCGGTGGGGTGATG 855
              ||| ||||   | |||||||||| || ||||||||| ||||||||||||  || ||||||
Db        472 GTGGTGTCACACTCGCTCTGCTCAGCTCTCTACGTGTGGGACCTCTGCGACGGAGTGATG 413
```

```
Qy        856 CTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTTGTCCAAGACTGCAAT 915
              || |||| ||  || |||   ||| ||| ||||||||   |||||| || || |||||
Db        412 CTTGCAGTTCAGCTGATCA---TCTGGCCTCAGCACCATGAGTTTGTGCATGAATGCAAC 356

Qy        916 TGCTCCATCTACCCTGGTACCATCACTGGACACCGCATG-GCATGGGACATGATGATGAA 974
              |||||||||| ||||| ||||||||||||||| ||| ||||||||||||||||||||
Db        355 TGCTCCATCTATCCTGGCGCCATCACTGGACACCGTATGAGCATGGGACATGATGATGAA 296

Qy        975 CTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTCCCCGAGGTCATTAT 1034
              ||||||| ||| ||| | |||||| |||||||||| ||||| || || |||||||| ||
Db        295 CTGGTCGTGCACCGCTGCTATGATCATGGCGTACGCAATGCGCGTTCCTGAGGTCATCAT 236

Qy       1035 AGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCCTACTTCTCTATGCA 1094
              ||| ||||| |||||||||||| |||||||||||||||||||| ||||||||||||||||
Db        235 AGATATCATCAGCGGGGCTCACTGGGGCGTCATGTTCGGCTTAGCCTACTTCTCTATGCA 176

Qy       1095 GGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGGGTGGACGCGCGCAC 1154
              |||||||||||||||||| ||||||||||||||| |||| ||||| ||||| |||| ||||
Db        175 GGGAGCGTGGGCGAAAGTCGCTGTCATCCTTCTGTTGACCGCTGGGGTACACGCGAGCAC 116

Qy       1155 CCATACTGTTGGGGGTTCTGC-CGCGCAGACCACCGGGCGCCTCACCAGCTTATTTGACA 1213
              |||       || ||  |||| ||||   || || |  |  ||  |||||  |||  |
Db        115 CCACGTCACAGGTGGCACTGCTAGCGCGCGCCGCCAGCAGTTTCGTCAGCTTGCTCACCC 56

Qy       1214 TGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1260
                ||| | |  ||||| ||||| ||| |||  | | || |||   ||
Db         55 CTGGCGCTAAACAGAACATCCAACTCATTANGATCNATTGCAATGGG 9
```

RESULT 2
AV758366/c
LOCUS AV758366 492 bp mRNA linear EST 19-OCT-2000
DEFINITION AV758366 BM Homo sapiens cDNA clone BMFAKA03 5', mRNA sequence.
ACCESSION AV758366
VERSION AV758366.1 GI:10916214
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 492)
AUTHORS Gu,J., Zhao,M., Huang,Q., Xu,X., Li,Y., Peng,Y., Song,H., Xiao,H., Gu,Y., Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Yang,Y., Gao,G., Wang,Z., Zhang,Q., Chen,S., Han,Z. and Chen,Z.
TITLE Homo sapiens cDNA BM clones
JOURNAL Unpublished (2000)
COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES Location/Qualifiers
source 1. .492
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="BMFAKA03"
/tissue_type="Bone marrow"
/cell_type="CD34+ hematopoietic stem/progenitor cell"
/lab_host="BM25.8"
/clone_lib="BM"
/note="Vector: pTriplEx2; Site_1: sfiIA; Site_2: sfiIB"
ORIGIN

Query Match 2.6%; Score 235.8; DB 1; Length 492;
Best Local Similarity 72.0%; Pred. No. 1.7e-53;
Matches 347; Conservative 0; Mismatches 130; Indels 5; Gaps 3;

```
Qy        783 CATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTCTG 842
              || | |||| ||||||     |   || ||| || || || | |||||||  |||||||||
Db        486 CACCTACATAGTTGTGCATACACACATGCTGTGATCAGCTCACTACGTGTTGGACCTCTG 427

Qy        843 CGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTTGT 902
              || ||||||   ||| |||||||| ||    ||| |||| | |||||| || ||||||||
Db        426 CGTTGGGGTATCGCTTGCAGCCCAACTG---ATTATCTCTCAGCAGCAACATTGGTTTGT 370

Qy        903 CCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGAC-ACCGCATGGCATGGG 961
              |||||  ||||| |||||  |||| |||||   |||||||||| || | ||||||| ||
Db        369 GCAAGAATGCAACTGCTCATTCTATCCTGGCTGCATCACTGGACTACAGTATGGCATAGG 310

Qy        962 ACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTCC 1021
              |||||||||||||||||| |  | | | |||||||  |||||||||| ||||| || |
Db        309 CTATGATGATGAACTGGTCGGCAGCCGGTTCCATGATACTGGCGTACGCAATGCGCGTTC 250

Qy       1022 CCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCCT 1081
              | || ||| | ||||| ||||| ||| |||| || || ||||||||||||||||  || |
Db        249 CTGAAGTCCTCATAGATATCATAAGCTGGGCACACTGCGGCGTCATGTTCGGCTCAGCTT 190

Qy       1082 ACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGGG 1141
              ||||| | |||||||||||||| ||| ||||| | |||||||| ||||||  |||| ||||
Db        189 ACTTCACAATGCAGGGAGCGTTGGCCAAAGTAGCTGTCATCCATCTGTTCACCGCTGGGG 130

Qy       1142 TGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACCA 1201
              | || ||| ||| |||        || |||     | ||||     |  | |  |  ||  ||
Db        129 TTGATGCGAGCAACCACGTATCAGGTGGTATCACTGCGCGCGACGACAGCAGTTTCGTCA 70

Qy       1202 GCTTATTTGA-CATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1260
              | ||   | |     ||| | | |||| | |||||||| | | |  ||||||||   ||
Db         69 GGTTGCATAATGCCTGGCGCTAAGCAGTACATCCAGCTCATGANCGTCAATGGCAATNGG 10

Qy       1261 CA 1262
                |
Db           9 NA 8
```

RESULT 3
R28798
LOCUS R28798 179 bp mRNA linear EST 03-DEC-1999
DEFINITION F0-266D 22 week old human fetal liver cDNA library Homo sapiens cDNA clone F0-266D 5', mRNA sequence.
ACCESSION R28798
VERSION R28798.1 GI:6514170
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 179)
AUTHORS Choi,S.S., Yun,J.W., Choi,E.K., Cho,Y.G., Sung,Y.C. and Shin,H.S.
TITLE Construction of a gene expression profile of a human fetal liver by single-pass cDNA sequencing
JOURNAL Mamm. Genome 6 (9), 653-657 (1995)
MEDLINE 96081342
PUBMED 8535075
COMMENT Contact: Hee-Sup Shin
Developmental Genetics
Pohang Institute of Science & Technology
San31, Hyojadong Pohang, 790-784 Republic of Korea
Tel: 562-279-2291
Fax: 562-279-2199
Email: shinhs@vision.postech.ac.kr
Seq primer: T3 primer.
FEATURES Location/Qualifiers
source 1. .179
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="F0-266D"
/lab_host="XL1-blue MRF'"
/clone_lib="22 week old human fetal liver cDNA library"

```
                     /note="Vector: pBluescriptII SK(-); Site_1: EcoRI; Site_2:
                     XhoI; The cDNA library made by oligo-dT primed and
                     directionally cloned between 5'ExoR I-XhoI3' sites."
ORIGIN

  Query Match             0.6%;  Score 57.6;  DB 7;  Length 179;
  Best Local Similarity   78.4%;  Pred. No. 0.00056;
  Matches   69;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy       1210 GACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGGCACATCAAC 1269
             ||||| |       |||||| || |||||||| ||| | ||||||||||||||||||||
Db         50 GACATTGTAGATCCGCAGAACATTCAGCTCGTAAACGCTAATGGCAGCTGGCACATCAAC 109

Qy       1270 CGCACCGCCCTGAACTGCAATGACTCCT 1297
              | || ||| |||| |||||||| ||||
Db        110 AGGACTGCCTTGAATTGCAATGAGTCCT 137


RESULT 4
CN373175
LOCUS       CN373175                 522 bp    mRNA    linear   EST 16-MAY-2004
DEFINITION  17000531390153 GRN_EB Homo sapiens cDNA 5', mRNA sequence.
ACCESSION   CN373175
VERSION     CN373175.1  GI:47373109
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 522)
  AUTHORS   Brandenberger,R., Wei,H., Zhang,S., Lei,S., Murage,J., Fisk,G.J.,
            Li,Y., Xu,C., Fang,R., Guegler,K., Rao,M.S., Mandalam,R.,
            Lebkowski,J and Stanton,L.W.
  TITLE     Transcriptome characterization elucidates signaling networks that
            control human ES cell growth and differentiation
  JOURNAL   Nat. Biotechnol. 22 (6), 707-716 (2004)
COMMENT     Contact: Brandenberger R
            Regenerative Medicine
            Geron Corporation
            230 Constitution Drive, Menlo Park, CA 94025, USA
            Tel: 650 473 8658
            Fax: 650 473 7760
            Email: rbrandenberger@geron.com
            Insert Length: 522   Std Error: 0.00.
FEATURES             Location/Qualifiers
     source          1. .522
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="embryonic stem cells, embryoid bodies
                     derived from H1, H7 and H9 cells"
                     /clone_lib="GRN_EB"
                     /note="oligo dT primed, full-length enriched cDNA library
                     from embryoid body outgrowths derived from hES cell lines
                     H1 (p32), H7 (p29), and H9 (p26) maintained in feeder-free
                     conditions."
ORIGIN

  Query Match             0.6%;  Score 52;  DB 7;  Length 522;
  Best Local Similarity   46.8%;  Pred. No. 0.029;
  Matches  163;  Conservative    0;  Mismatches  185;  Indels    0;  Gaps    0;

Qy        748 CGGCCCGGCGCCCTCACGCAGGGCTTGCGGACGCACATCGACATGGTTGTGATGTCCGCC 807
              | ||||   | |  | | | |  ||  | |        | |  |||  ||| |    |
Db         59 CAGCCCAATGTCAGCGCCCTGCACTGCCAGCACTTGGGCAATGTGGCCGTGCTCATGGAG 118

Qy        808 ACGCTCTGCTCTGCCCTCTACGTGGGGGACCTCTGCGGTGGGGTGATGCTCGCAGCCCAA 867
              |  |  || | ||     | |||||||  |  | | |  ||| ||    | ||  |   |
Db        119 CTGAGCGCCTTTCCCAGGGAGGTGGGGGGCGCCGGGGCAGGGCTGCACCCCGTGGTATAC 178

Qy        868 ATGTTCATTGTCTCGCCGCAGCACCACTGGTTTGTCCAAGACTGCAATTGCTCCATCTAC 927
              | ||  | || || || || | |   || | |     |  ||  | |  | || ||
Db        179 CCCTGCACGGCCTTGCTGCTGCTCTGCCTCTTCGCCACCATCATCACCTACATCCTCAAC 238

Qy        928 CCTGGTACCATCACTGGACACCGCATGGCATGGGACATGATGATGAACTGGTCGCCCACG 987
              |  |  |||||  ||    ||| |  |  ||| ||||| || |||||| ||    |
Db        239 CACAGCTCCATCCGTGTGTCCCGGAAAGGCTGGCACATGCTGCTGAACTTGTGCTTCCAC 298

Qy        988 GCTACCATGATCTTGGCGTACGCGATGCGTGTCCCCGAGGTCATTATAGACATCATTAGC 1047
              ||||||  ||  ||   |    || | | |  |    |||  |   ||   ||   |
Db        299 ATAGCCATGACCTCTGCTGTCTTTGTGGGGGGCATCACACTCACCAACTACCAGATGGTC 358

Qy       1048 GGGGCTCATTGGGGCGTCATGTTCGGCTTGGCCTACTTCTCTATGCAG 1095
              |          |||| |||    |   ||   ||| | | || | || |
Db        359 TGCCAGGCGGTGGGCATCACCCTGCACTACTCCTCCCTATCCACGCTG 406


RESULT 5
CNS0091P
LOCUS       CNS0091P                 925 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL053013
VERSION     AL053013.1  GI:4934461
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 925)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .925
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR19D16"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             0.6%;  Score 52;  DB 9;  Length 925;
  Best Local Similarity   12.8%;  Pred. No. 0.035;
  Matches   46;  Conservative  167;  Mismatches  147;  Indels    0;  Gaps    0;

Qy        668 TCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATACCGGTCT 727
              |:: :| |:||:  : :  :  :     :::    :|  | : :  | :: : :::  :|
Db        554 TSSGSGYGKGCSSGSGBSCSCCSSCSCSSSSCSCCBCCCCCSCSSYCCSSSBSSSKCSST 613

Qy        728 CACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACGCACATCG 787
              : :|:   : :::|||:| ::::|::::::  :: : ::: ::::  |::: ::: : |::
Db        614 SBSCSCCCSSKSVCGTSCSSSSSCSSSSSSTSSSSTSSSTSKSSSSGSSSSSSSSYTTSK 673

Qy        788 ACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTCTGCGGTG 847
                :|  :::| | : :    :: : ::    ::: :   :::: ::  ::: : :: :
```

```
Db        674 STSASGSGSWSAGGGSGSTGSTSSSSSSSSTSTSSSSVSSGSKSSTBSSGSBSSSGSSSS 733

Qy        848 GGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTTGTCCAAG 907
              ::  : ::||: : ::::    :| :  |  | :::::: :: :: ::|:|  : :: :|
Db        734 SSTSSBBSCTSTSSSSSSSYSSSTCSCCTCCCSYSYSSSTSSSSSTSWGSTSGSSSSSVG 793

Qy        908 ACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGGGACATGA 967
              : ::: |  | :|  :   ||   :::| |: : || : ::|   :     ||    :
Db        794 TSSSSDSTSTCCSCCCYMCTCCSTYBMBCYTSTSCGGSSSSSGKGGVTKCGCGGCGSSST 853

Qy        968 TGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTCCCCGAGG 1027
              |::|  :  : ::::| :::: ::  :| :  ::::  :|::  ::: ::::::| ::
Db        854 NGMBGTSSACSSSSSSCSSSSVSSSSKSSASSSSVSSSGSSGVSSNSSSASKSSSSGSVS 913


RESULT 6
CNS0091P/c
LOCUS       CNS0091P                925 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL053013
VERSION     AL053013.1  GI:4934461
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 925)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .925
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR19D16"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             0.6%;  Score 51.8;  DB 9;  Length 925;
  Best Local Similarity   15.6%;  Pred. No. 0.04;
  Matches   73;  Conservative  178;  Mismatches  211;  Indels    5;  Gaps    1;

Qy       6935 CGGGCTGTGCTCTCCCTCCTCCTAGGAAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGA 6994
              |:::|    : : ::| :: ::| ::    |::|: :::: : :::  :: :::: :::
Db        925 CSBSCSCSCSCSSBSCSSSSMSTSSSNSSBCSSCSSSBSSSSTSSMSSSSBSSSSGSSSS 866

Qy       6995 CAGTGGGCCTAAGTGAGGACTCCATAGGAGATGCCCTTCAACAGCTGGCCATTAAGTCCT 7054
              : ||:: |::    :  | | | :    :  :::|  :|  :|:::: :     :
Db        865 SSGTSSACVKCNASSSCGCCGCGMABCCMCSSSSSCCGSASARGVKVRASGGAGKRGGGS 806

Qy       7055 TTGGCCAGCCCCCCCCAAGCGGCGATTCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATT 7114
```

```
                : :::::: |:::  :|: : :  : :::     :::||| | :| ::: ::::
Db        805 GGASASHSSSSACBSSSSSCSASCWSASSSSSASSSRSRSGGGAGGSGASSSRSSSSSSS 746

Qy       7115 CCGGCAGTCAGACGCCTCCTGATGAGTTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCA 7174
              : |::: :|: :::: :: :  :  :  : :|: : :::|:|: ::  ::| | :  ||
Db        745 ASAGSVVSSASSSSSSCSSSVSCSSVASSMSCSSBSSSSASASSSSSSSSSASCASCSCCC 686

Qy       7175 TGCCCCC-----CCTCGAGGGGGAGCTTGGAGATCCAGACCTGGAGCCTGAGCAGGTAGA 7229
              |: :|:|     ::  : ::::: :|  ::::| ::|: :: :: ::: : :: |:  |
Db        685 TSWSCSCSTSASMSAARSSSSSSSSCSSSSMSASSSASSSSASSSSSSGSSSSSGSACGB 626

Qy       7230 GCCCCAACCCCCCCCCCAGGGGGGGGTGGCAGCTCCCGGCTCGGACTCGGGGTCCTGGTC 7289
              ::::   : ::: :: : ::::::|| ::  |    :||: :::     ::||  :  : :
Db        625 SMSSGGGSGSVSASSGMSSSVSSSGGRSSGSGGGGGVGGSGSSSSSGSGSSGGSGSVCSCS 566

Qy       7290 TACTTGCTCCGAGGAGGACGACTCCGTCGTGTGCTGCTCCATGTCATACTCCTGGACCGG 7349
                |  :| |::|  |   |:|  |   ||:::         | |     |  |
Db        565 SGCMCRCSCSSAAAAAASCVAASCGMMCGKSKSGCTNNNNNTNTNTNTNNTNNNNTTA 506

Qy       7350 GGCTCTAATAACTCCTTGTAGTCCCGAAGAGGAGAAGTTACCGATTA 7396
                |||       ||             | |    |||   :|||
Db        505 ATAAATAANNNTNTATTNNNNNTNNNNNNANAAANTATTATTTWTTA 459


RESULT 7
CNS015Y4/c
LOCUS       CNS015Y4                1203 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN15E10 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL106054
VERSION     AL106054.1  GI:5619805
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1203)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1203
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN15E10"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : T7"
ORIGIN

  Query Match             0.6%;  Score 51.4;  DB 9;  Length 1203;
  Best Local Similarity   34.6%;  Pred. No. 0.057;
  Matches  128;  Conservative   69;  Mismatches  171;  Indels    2;  Gaps    1;

Qy       3322 GCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGGACTAAATCTTTGGAGCCGTGC 3381
              || | |||||     |  :|| | ||: |: :|:| |:| :     :   || |    ||
Db       1198 GCGGGGGGGGGGGGGGGCSGGSGCCCSCGSGSCSCSGSGGSCCSCGSGGSGGGGGGCGGC 1139

Qy       3382 ACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAACGCTGATGTCATCCCGGCTCGA 3441
              :| | |   ||| || ||    || ::    |: :   |: |       :: ||  |
```

```
Db       1138 CSGGGGGCCCCGGGCGGCCCCCCCCSSSGGSGGSCSGGGGSGGGCCGGCSSSGGGGSCCG 1079

Qy       3442 AGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGACCTCTTTCCACCTTGAAGGGG 3501
              | :| |::  :   ::||| | |  | | |||| : : || : : || ||: :  ::
Db       1078 GGGSGGGSSCCSSSSSSGGGGCCCCCCCCCCCCCCSSSSCCCSSKCCCCCCKNSKCSKCC 1019

Qy       3502 TCCTCAGGAGGCCCGG--TGCTATGCCCCAGGGGCCACGCTGTCGGAGTCTTCCGGGCAG 3559
              || | :| | |||    ::|   :|||| :: || |:| : |:: : :::||:||:
Db       1018 CCCSCSSGGGCCCCCCCCKBCGCCBCCCCCCSSCCCCCSCSSCCSBSSSSBBCCKGGSKC 959

Qy       3560 CTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATCCCCGTTGAGACACTCGACA 3619
              |   :|:| ::|  ::|  :::   :::|      ::|  ||||       | | |  |
Db        958 CCCCBTKCCBKCCCKSCCCBKBCCCBBBCCCSCCSBKCCCCCCCNCCCCCCCCCCCCCCS 899

Qy       3620 TCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCACCTGCTGTGCCCCAGACCT 3679
              |  : :|||: ||||| |:    | | |  |  | | |  |: |  | :|||: | ||
Db        898 CCCCSSSGCGSSCCCCCCCSGGGSGGGNCCCCCCCCCCCGGCSGGGGGGSCCCSGGGCCC 839

Qy       3680 ATCAGGTCGG 3689
              |  : :||
Db        838 CCCCCSGSGG 829
```

```
RESULT 8
HSM804270
LOCUS       HSM804270              4762 bp    mRNA    linear   HTC 22-SEP-2004
DEFINITION  Homo sapiens mRNA; cDNA DKFZp666M0710 (from clone DKFZp666M0710).
ACCESSION   AL832959
VERSION     AL832959.1  GI:21733547
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4762)
  AUTHORS   Ansorge,W., Krieger,S., Regiert,T., Rittmueller,C., Schwager,B.,
            Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
            Wiemann,S.
  CONSRTM   The German cDNA Consortium
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-2004) MIPS, Ingolstaedter Landstr.1, D-85764
            Neuherberg, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by EMBL (European Molecular Biology Laboratories,
            Heidelberg/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            This clone (DKFZp666M0710) is available at the RZPD Deutsches
            Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
            Please contact RZPD for ordering:
            http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp666M0710
            Further information about the clone and the sequencing project is
            available at http://mips.gsf.de/projects/cdna/.
FEATURES             Location/Qualifiers
     source          1. .4762
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="RZPD:DKFZp666M0710"
                     /db_xref="taxon:9606"
                     /clone="DKFZp666M0710"
                     /tissue_type="stomach"
                     /clone_lib="666 (synonym: hsto2). Vector pSport1; host
                     DH10B; sites NotI + SalI"
                     /dev_stage="adult"
                     /note="G protein-coupled receptor 124 precursor,
                     N-terminus truncated"
     gene            1. .4762
                     /gene="DKFZp666M0710"
     CDS             <2. .2719
                     /gene="DKFZp666M0710"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAH56283.1"
                     /db_xref="GI:52545734"
                     /translation="SHIGVWASGEWECTVSMAQGNASKKVEIVVLETSASYCPAERVA
                     NNRGDFRWPRTLAGITAYQSCLQYPFTSVPLGGGAPGTRASRRCDRAGRWEPGDYSHC
                     LYTNDITRVLYTFVLMPINASNALTLAHQLRVYTAEAASFSDMMDVVYVAQMIQKFLG
                     YVDQIKELVEVMVDMASNLMLVDEHLLWLAQREDKACSRIVGALERIGGAALSPHAQH
                     ISVNARNVALEAYLIKPHSYVGLTCTAFQRREGGVPGTRPGSPGQNPPPEPEPPADQQ
                     LRFRCTTGRPNVSLSSFHIKNSVALASIQLPPSLFSSLPAALAPPVPPDCTLQLLVFR
                     NGRLFHSHSNTSRPGAAGPGKRRGVATPVIFAGTSGCGVGNLTEPVAVSLRHWAEGAE
                     PVAAWWSQEGPGEAGGWTSEGCQLRSSQPNVSALHCQHLGNVAVLMELSAFPREVGGA
                     GAGLHPVVYPCTALLLLCLFATIITYILNHSSIRVSRKGWHMLLNLCFHIAMTSAVFA
                     GGITLTNYQMVCQAVGITLHYSSLSTLLWMGVKARVLHKELTWRAPPPQEGDPALPTP
                     SPMLRFYLIAGGIPLIICGITAAVNIHNYRDHSPYCWLVWRPSLGAFYIPVALILLIT
                     WIYFLCAGLRLRGPLAQNPKAGNSRASLEAGEELRGSTRLRGSGPLLSDSGSLLATGS
                     ARVGTPGPPEDGDSLYSPGVQLRGAGDHALPVLGHVGLRGSGSVPALAAPGGVQLLVR
                     GGSLRPGPLRLHSPLCQAEGRESLVARLLPPCLSRGPPCPAPGPARRRRGRFPGVRGG
                     APLPQVLPKRQQRPSAGSGPLQAHQPAAGPESGVRGGGGGRRGRRAGAGGHPGKPRPP
                     PPQQRAPRASGAQEPGQGTPRGGGLRQEPAQGPARGRGGGAGAAVQRER"
ORIGIN

  Query Match             0.6%;  Score 50.6;  DB 3;  Length 4762;
  Best Local Similarity   42.6%;  Pred. No. 0.15;
  Matches  263;  Conservative    0;  Mismatches  354;  Indels    0;  Gaps    0;

Qy        684 GTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATACCGGTCTCACCGAATGTGGCCGT 743
              | || || | | |       |   ||  | ||||||   |||         |  |   |
Db       1195 GAGCCAGGAGGGGCCCGGGGAGGCTGGGGGCTGGACCTCGGAGGGCTGCCAGCTCCGCTC 1254

Qy        744 GCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACGCACATCGACATGGTTGTGATGTC 803
              | ||||  | | | | | | ||  | |          | |  |||  ||| |
Db       1255 CAGCCAGCCCAATGTCAGCGCCCTGCACTGCCAGCACTTGGGCAATGTGGCCGTGCTCAT 1314

Qy        804 CGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTCTGCGGTGGGGTGATGCTCGCAGC 863
              |    |  |  || | ||    | |||||| |  | | |  ||| ||   | ||  |
Db       1315 GGAGCTGAGCGCCTTTCCCAGGGAGGTGGGGGGCGCCGGGGCAGGGCTGCACCCCGTGGT 1374

Qy        864 CCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTTGTCCAAGACTGCAATTGCTCCAT 923
              |    | ||  | || || || || |  |   || | |     |  ||  | |  | |
Db       1375 ATACCCCTGCACGGCCTTGCTGCTGCTCTGCCTCTTCGCCACCATCATCACCTACATCCT 1434

Qy        924 CTACCCTGGTACCATCACTGGACACCGCATGGCATGGGACATGATGATGAACTGGTCGCC 983
              | |||    |  ||||| ||    ||| |  |  ||| ||||| || |||||| ||
Db       1435 CAACCACAGCTCCATCCGTGTGTCCCGGAAAGGCTGGCACATGCTGCTGAACTTGTGCTT 1494

Qy        984 CACGGCTACCATGATCTTGGCGTACGCGATGCGTGTCCCCGAGGTCATTATAGACATCAT 1043
              |      |||||| ||  ||   |    | | | |  |    |||  |   ||  ||
Db       1495 CCACATAGCCATGACCTCTGCTGTCTTTGCGGGGGGCATCACACTCACCAACTACCAGAT 1554

Qy       1044 TAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCCTACTTCTCTATGCAGGGAGCGTG 1103
              | |         |||| |||   |  ||  ||| | | || | || |     |
Db       1555 GGTCTGCCAGGCGGTGGGCATCACCCTGCACTACTCCTCCCTATCCACGCTGCTCTGGAT 1614

Qy       1104 GGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGGGTGGACGCGCGCACCCATACTGT 1163
              ||        |  || |    ||    |    |  || | | | |     |
Db       1615 GGGCGTGAAGGCGCGAGTGCTCCATAAGGAGCTCACCTGGAGGGCACCCCCTCCGCAAGA 1674

Qy       1164 TGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACCAGCTTATTTGACATGGGCCCCAG 1223
              ||||  | ||   ||  ||  || | |   |  | || | |  || | |       |
Db       1675 AGGGGACCCCGCTCTGCCTACTCCCAGTCCTATGCTCCGGTTCTATTTGATCGCTGGAGG 1734

Qy       1224 GCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGGCACATCAACCGCACCGCCCTGAA 1283
              |     | ||    ||  |  || ||  |         ||||| ||  |  |       |
Db       1735 GATTCCACTCATTATCTGTGGCATCACAGCTGCAGTCAACATCCACAACTACCGGGACCA 1794

Qy       1284 CTGCAATGACTCCTTGC 1300
              | ||    ||| || ||
Db       1795 CAGCCCCTACTGCTGGC 1811
```

```
RESULT 9
AY411015
LOCUS       AY411015               3661 bp    DNA     linear   GSS 16-DEC-2003
DEFINITION  Homo sapiens GPR124 gene, VIRTUAL TRANSCRIPT, partial sequence,
            genomic survey sequence.
```

```
ACCESSION   AY411015
VERSION     AY411015.1  GI:39766983
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3661)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE     Inferring nonneutral evolution from human-chimp-mouse orthologous
            gene trios
  JOURNAL   Science 302 (5652), 1960-1963 (2003)
   PUBMED   14671302
REFERENCE   2  (bases 1 to 3661)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
COMMENT     This sequence was made by sequencing genomic exons and ordering
            them based on alignment.
FEATURES             Location/Qualifiers
     source          1. .3661
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     gene            <1. .>3661
                     /gene="GPR124"
                     /locus_tag="HCM4083"
ORIGIN

  Query Match             0.5%;  Score 50;  DB 9;  Length 3661;
  Best Local Similarity   16.7%;  Pred. No. 0.2;
  Matches  170;  Conservative    0;  Mismatches  851;  Indels    0;  Gaps    0;

Qy         75 GGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGGGGCCCCAGGTTGGG 134
              || | || |||  ||||| || |   |  |     |   | |    | || |||
Db       1264 GGCCTGCAGCCGCATCGTGGGTGCCCTGGAGCGCATTGGGGGGGCCGCCCTCAGCCCCCA 1323

Qy        135 TGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGAAGGCGCCAGCCCAT 194
              ||  |     |      |
Db       1324 TGCCCAGCACATCTCAGTGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1383

Qy        195 CCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGATACCCCTGGCCCCT 254

Db       1384 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1443

Qy        255 ATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCCCGAGGTTCCCGTCC 314

Db       1444 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1503

Qy        315 CTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGTAAGGTCATCGATAC 374

Db       1504 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1563

Qy        375 CCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTGGGCGCCCCGCTCGG 434

Db       1564 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1623

Qy        435 CGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGACGGGGTTAATTTTGC 494

Db       1624 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1683

Qy        495 AACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCCCTGCTGTCCTGCAT 554

Db       1684 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1743

Qy        555 CACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGCTACATGGTGACTAA 614

Db       1744 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1803

Qy        615 CGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTCCTCCACGTCCCCGG 674

Db       1804 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1863

Qy        675 GTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATACCGGTCTCACCGAA 734

Db       1864 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1923

Qy        735 TGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACGCACATCGACATGGT 794

Db       1924 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1983

Qy        795 TGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTCTGCGGTGGGGTGAT 854
                  |    |  |  || | ||    | |||||||  |  | | |  ||| ||
Db       1984 NNNNNNNNNGAGCTGAGCGCCTTTCCCAGGGAGGTGGGGGGCGCCGGGGCAGGGCTGCA 2043

Qy        855 GCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTTGTCCAAGACTGCAA 914
              | ||  |   |    | ||  | || || || || |  |   || | |     |  ||
Db       2044 CCCCGTGGTATACCCCTGCACGGCCTTGCTGCTGCTCTGCCTCTTCGCCACCATCATCAC 2103

Qy        915 TTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGGGACATGATGATGAA 974
              | |  | || |||    |  |||||  ||    ||| |  |  ||| ||||| || ||||
Db       2104 CTACATCCTCAACCACAGCTCCATCCGTGTGTCCCGGAAAGGCTGGCACATGCTGCTGAA 2163

Qy        975 CTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTCCCCGAGGTCATTAT 1034
              || ||    |       |||||| ||  ||  |      | | | |  |    ||| |
Db       2164 CTTGTGCTTCCACATAGCCATGACCTCTGCTGTCTTTGCGGGGGGCATCACACTCACCAA 2223

Qy       1035 AGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCCTACTTCTCTATGCA 1094
                ||   ||   | |          |||| |||   |   ||   ||| | | || | ||
Db       2224 CTACCAGATGGTCTGCCAGGCGGTGGGCATCACCCTGCACTACTCCTCCCTATCCACGCT 2283

Qy       1095 G 1095
                |
Db       2284 G 2284


RESULT 10
CNS0052P/c
LOCUS       CNS0052P                844 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR11P16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL056652
VERSION     AL056652.1  GI:4932342
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 844)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
```

```
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .844
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR11P16"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             0.5%;  Score 49.2;  DB 9;  Length 844;
  Best Local Similarity   16.5%;  Pred. No. 0.21;
  Matches   67;  Conservative  165;  Mismatches  174;  Indels    0;  Gaps    0;

Qy       6904 AAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTCCCTCCTCCTAGGAAA 6963
Db        801 ARMAAAAAAAMDAACCCCCCCCCCCMCMCHMCMCCMMVCCTTTTHTTTYTCMYWAMMRMY 742

Qy       6964 ACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGTGAGGACTCCATAGGA 7023
Db        741 AWMMMMAMMWVTMCVMMMAVAMVMYAMAMAMRMNNVYCSBSCNBBTTTYYCNYTCVSSSB 682

Qy       7024 GATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCCCCAAGCGGCGATTCA 7083
Db        681 SSSBTTCWTHTNSNTYTMVTYVTTTHHTYTNTCCYYCCCCCCYYCSCCBYBCNTSCSYYS 622

Qy       7084 GGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACGCCTCCTGATGAGTTG 7143
Db        621 YSBCSGCSSSSSSSCSSGSSBYSCCCBCCCBGYCGCYSBVCSSBSSTBSSVSGSBSSTSS 562

Qy       7144 GCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCTCGAGGGGGAGCTTGGAGAT 7203
Db        561 GTGCKCBSSSSSBBCSSBSSCCCYCSSYCTYCYSSCCBYCCBSGCBTBSYBSCSSGBTGB 502

Qy       7204 CCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCAGGGGGGGGTGGCAGCT 7263
Db        501 SSBSGGKSGGKSSGKGGGBKSGGKGYCYCBTYCCSBSSBCCSSSSSSTSKCCSGGSSSTK 442

Qy       7264 CCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGACG 7309
Db        441 CGSSBBGGGSSSKTSGSSSCBSGSSSTGSSGSSKKYSYSTSSSSSG 396


RESULT 11
BQ940004
LOCUS       BQ940004                1082 bp    mRNA    linear   EST 21-AUG-2002
DEFINITION  AGENCOURT_8854250 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:6473512
            5', mRNA sequence.
ACCESSION   BQ940004
VERSION     BQ940004.1  GI:22355482
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1082)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM14009  row: c  column: 17
            High quality sequence stop: 460.
FEATURES             Location/Qualifiers
     source          1. .1082
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6473512"
                     /tissue_type="leiomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_71"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 2.1 kb.  "
ORIGIN

  Query Match             0.5%;  Score 48.8;  DB 5;  Length 1082;
  Best Local Similarity   44.9%;  Pred. No. 0.29;
  Matches  185;  Conservative    0;  Mismatches  227;  Indels    0;  Gaps    0;

Qy        684 GTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATACCGGTCTCACCGAATGTGGCCGT 743
Db         26 GAGCCAGGAGGGGCCCGGGGAGGCTGGGGGCTGGACCTCGGAGGGCTGCCAGCTCCGCTC 85

Qy        744 GCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACGCACATCGACATGGTTGTGATGTC 803
Db         86 CAGCCAGCCCAATGTCAGCGCCCTGCACTGCCAGCACTTGGGCAATGTGGCCGTGCTCAT 145

Qy        804 CGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTCTGCGGTGGGGTGATGCTCGCAGC 863
Db        146 GGAGCTGAGCGCCTTTCCCACGGAGGTGGGGGGCGCCGGGGCAGGGCTGCACCCCGTGGT 205

Qy        864 CCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTTGTCCAAGACTGCAATTGCTCCAT 923
Db        206 ATACCCCTGCACGGCCTTGCTGCTGCTCTCTGCCTCTTCGCCACCATCATCACCTATATCCT 265

Qy        924 CTACCCTGGTACCATCACTGGACACCGCATGGCATGGGACATGATGATGAACTGGTCGCC 983
Db        266 CAACCACAGCTCCATCCGTGTGTCCCGGAAAGGCTGGCACATGCTGCTGAACTTGTGCTT 325

Qy        984 CACGGCTACCATGATCTTGGCGTACGCGATGCGTGTCCCCGAGGTCATTATAGACATCAT 1043
Db        326 CCACATAGCCATGACCTCTGCTGTCTTTGCGGGGGGCATCACACTCACCAACTACCAGAT 385

Qy       1044 TAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCCTACTTCTCTATGCAG 1095
Db        386 GGTCTGCCAGGCGGTGGGCATCACCCTGCACTACTCCTCCCTATCCACGCTG 437


RESULT 12
BH622015/c
LOCUS       BH622015                 393 bp    DNA     linear   GSS 30-JAN-2002
DEFINITION  1007116B04.2EL_y1 1007 - RescueMu Grid H Zea mays genomic, genomic
            survey sequence.
ACCESSION   BH622015
VERSION     BH622015.1  GI:18435266
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 393)
  AUTHORS   Walbot,V.
  TITLE     Maize genomic sequences found using engineered RescueMu transposon
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Possible ligation site of ends cut by 2 different endonucleases.
            Reverse complemented post-ligation sequence from source sequence.
            Plate: 1007116  column: 24
```

```
          Class: transposon-tagged.
FEATURES             Location/Qualifiers
     source          1. .393
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="mixed background W23/A188/B73"
                     /db_xref="taxon:4577"
                     /tissue_type="leaf"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="1007 - RescueMu Grid H"
                     /note="Organ: leaf; Vector: RescueMu (engineered from
                     pBlueScript backbone); Site_1: BamHI; Site_2: BglII;
                     RescueMu is a 4.9 kb, modified maize Mu transposon
                     designed to allow plasmid rescue from total genomic DNA.
                     Mu elements insert preferentially into transcription
                     units.  For more information on RescueMu, go to the  web
                     site 'www.zmdb.iastate.edu' and follow the links for
                     'RescueMu.'  Grid H was grown at Berkeley in 2001.  DNA
                     was extracted from leaf punches,  double digested using
                     BamHI and BglII, and ligated to form circular plasmids.
                     DH10B cells were transformed and then screened on LB
                     plates with ampicillin."
ORIGIN

  Query Match             0.5%;  Score 47.2;  DB 8;  Length 393;
  Best Local Similarity   47.1%;  Pred. No. 0.58;
  Matches  145;  Conservative    0;  Mismatches  163;  Indels    0;  Gaps    0;

Qy       1621 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 1680
              ||| | | |   ||| || | | |   | |  ||    | | || ||  ||
Db        334 TTGTGCTCCGCGGCACCGTCGGCGCCGGCCGCTTCGGCACGCCGCAGCAGGACCTAATCG 275

Qy       1681 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 1740
              ||| | |||  || | || | | |  ||  |||  |    || ||  | ||| |  |
Db        274 AGCTCCGCCAGCACCTTCGTCTCCCACCCGCACCGGGCGCTCGCCGAGCACCACGACGCC 215

Qy       1741 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 1800
              | |  |  |   | | ||||  |   |    |  |    |    || | | |  |
Db        214 AGGTCCAGGGCCTACACCACCACCACATCGTCGTCGTCGTCGTCCTCCTCCTCCGGC 155

Qy       1801 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 1860
              || ||   | ||||   |   |||| |||| || || |||   ||||||| |||  ||
Db        154 CGCCGCGACGAGCCCGACCCACCGCCGAGGTCGCTCATGGACCCGCCCTACACGCTGGGG 95

Qy       1861 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 1920
              |   |  ||||| ||    | || ||  |   | ||    ||  | | |   |
Db          94 CCCAAGATCTGCAGAGCCGCGTGTAACAAATAGTAGATTCTCTTGCCTCTTTGCTAGAGC 35

Qy       1921 GAGCACAG 1928
                |||| ||
Db          34 TAGCAAAG 27


RESULT 13
CNS010EW
LOCUS       CNS010EW                1009 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN03P19 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL098882
VERSION     AL098882.1  GI:5610493
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1009)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1009
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN03P19"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
ORIGIN

  Query Match             0.5%;  Score 46;  DB 9;  Length 1009;
  Best Local Similarity   31.6%;  Pred. No. 1.7;
  Matches   99;  Conservative   62;  Mismatches  152;  Indels    0;  Gaps    0;

Qy        145 ACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGAAGGCGCCAGCCCATCCCTAAAGAT 204
              ||||  ||||  :    ||   | |   |||:  : : | |  ::||: ||| : | |:
Db        510 ACAACAAAGAAAYACACGCACACACGCATACGBCTRNVACNCACVMCCMCCCCNMCACAY 569

Qy        205 CGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGATACCCCTGGCCCCTATACGGGAAT 264
              :: : | :| | | | |    ::::| | :|:: | |   :   : |      ||::
Db        570 SSCSCCCMCNCACGNACANARASRVRACACMCVRAASAGRAVAMMMACTAGSGCGSSGCG 629

Qy        265 GAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCCCGAGGTTCCCGTCCCTCTTGGGGC 324
              :|||  ::||  || | |  |:: :| : ||||:|:      |   ::| |      |
Db        630 TRGGGSRKSGGGGGGCCSCGGGGVSCSCCSCCCCCSCSCCCGCGCKCCMSCCCCCCCCCC 689

Qy        325 CCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGTAAGGTCATCGATACCCTAACGTGC 384
              |||   : : :::|||   || :: |    |:|||   |: |   :    ||   |: |:
Db        690 CCCGGCBGSGMSYGGCCGGGGGSSGGGGSGGKGGGGGGGSCCGCASCSGGCCGCGCSGGY 749

Qy        385 GGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTGGGCGCCCCGCTCGGCGGCGTCGCC 444
              : |   ||:  :| :  :  |  |  ::|    | | | || |::: |: |  :  |||
Db        750 SCCSSSGCSCSSCGSSSSCCGCSCCCSSCSCSNGCGCGTGCGCSSSCCSCCCCSCGCGCG 809

Qy        445 AGAGCTCTCGCGC 457
                |  | ||:::
Db        810 CCGGGSCGCGSSS 822


RESULT 14
CK307855/c
LOCUS       CK307855                755 bp    mRNA    linear   EST 01-MAR-2004
DEFINITION  SB02044B2D10.f1 normalized Keck-Tagu Library SB02 Taeniopygia
            guttata cDNA clone SB02044B2D10.f1 5, mRNA sequence.
ACCESSION   CK307855
VERSION     CK307855.1  GI:44817429
KEYWORDS    EST.
SOURCE      Taeniopygia guttata
  ORGANISM  Taeniopygia guttata
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Passeriformes; Estrildidae;
            Estrildinae; Taeniopygia.
REFERENCE   1  (bases 1 to 755)
  AUTHORS   Clayton,D.F., Arnold,A.P., Ball,G.F., Brenowitz,E., George,J.M.,
            Mello,C.V., Wade,J., Replogle,K., Lewin,H., Band,M., Hernandez,A.
            and Liu,L.
  TITLE     The Songbird Neurogenomics Initiative: An Evolving Public Resource
            for Study of Genes, Brain, and Behavior
  JOURNAL   Unpublished (2004)
COMMENT     Contact: David F. Clayton
            University of Illinois
            B107 CLSL, 601 S. Goodwin, Urbana, IL 61801, USA
            Tel: 217 244 3668
```

```
            Fax: 217 244 1648
            Email: dclayton@uiuc.edu
            Base Calling/Quality Scores: PHRED from Washington University
            Genome Center.
             Vector Trimming: Cross_match from Washington University Genome
            Center PHRAP suite. Low quality bases (Phred score < 20) were
            trimmed from both ends of the sequence by an in-house script.
             This sequence is vector free and at least 200 bp in length. Funded
            by PHS grant # RO1 NS045264, 'Songbird Neurogenomics Initiative.'
            PCR PRimers
            FORWARD: TAATACGACTCACTATAGGG(T7)
            BACKWARD: ATTAACCCTCACTAAAG(T3)
            Insert Length: 755   Std Error: 0.00
            Plate: SB02044B2  row: D  column: 10
            Seq primer: TAATACGACTCACTATAGGG (T7)
            High quality sequence stop: 755.
FEATURES              Location/Qualifiers
     source           1. .755
                      /organism="Taeniopygia guttata"
                      /mol_type="mRNA"
                      /db_xref="taxon:59729"
                      /clone="SB02044B2D10.f1"
                      /tissue_type="brain"
                      /dev_stage="late embryo, post-hatch days 1, 10, 20, 45,
                      and adult (pooled)"
                      /lab_host="DH10B"
                      /clone_lib="normalized Keck-Tagu Library SB02"
                      /note="Organ: brain; Vector: pBS II SK(+); Site_1:
                      EcoRI(5' side of insert); Site_2: NotI (3' side of
                      insert); The library was constructed and normalized as
                      described by Bonaldo, M.F., Lennon, G. and Soares, M.B.
                      (1996), Genome Research 6(9): 791-806. An identiying tag
                      was added at the 3'during cDNA synthesis:
                      insertAAAAAAAAAAAAAAAAAAATGCGA."
ORIGIN

  Query Match             0.5%;  Score 45.8;  DB 7;  Length 755;
  Best Local Similarity   48.6%;  Pred. No. 1.8;
  Matches  125;  Conservative    0;  Mismatches  132;  Indels    0;  Gaps    0;

Qy       1166 GGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACCAGCTTATTTGACATGGGCCCCAGGC 1225
             ||  | |||||  |  ||  | ||      | || |||||||  |     |    || ||
Db        353 GGATTGCTGCCTCATAGCGCTCCTTATTCTTCTCCAGCTTCTGCCTCTGCCGCTCCTCCA 294

Qy       1226 AGAAAATCCAGCTCGTTAACACCAATGGCAGCTGGCACATCAACCGCACCGCCCTGAACT 1285
              ||    || || | |   | ||    || ||||   |  || ||||     ||    ||
Db        293 GGACTGCCCGGCGCTTCTCCGCCCGCAGCCGCTGCTCCTCCAGCCGCTTGCGCCGCTCCT 234

Qy       1286 GCAATGACTCCTTGCACACCGGCTTTATCGCGTCTCTGTTCTACACCCACAGCTTCAACT 1345
              ||    || ||  | || | |       |    ||    |||| || |||||||  |  ||
Db        233 CCAGCTGCTTCTCCCGCAGCAGACGGGCCTTCTCCTCCTTCTCCAGCCACAGCACCCGCT 174

Qy       1346 CGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGCAGTATCGAGGCCTTCCGGGTGGGAT 1405
              | ||| | |   |      |   ||| |  |||||    |||     |||  |  | |  |
Db        173 TGGCAGCAAGGTACTTGGCCCGCTCCTCGCGCCGCTCCTTCGCCTGCTTGTGCCTTGCCT 114

Qy       1406 GGGGCGCCTTGCAATAT 1422
              ||| |  || |  || |
Db        113 GGGCCTTCTGGGCATCT 97


RESULT 15
CA733786
LOCUS       CA733786                562 bp    mRNA    linear   EST 26-NOV-2002
DEFINITION  wlp1c.pk006.d15 wlp1c Triticum aestivum cDNA clone wlp1c.pk006.d15
            5' end, mRNA sequence.
ACCESSION   CA733786
VERSION     CA733786.1  GI:25549384
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 562)
  AUTHORS   Tingey,S.V., Powell,W., Wolters,P., Dolan,M., Hainey,C., Yuan,Z.,
            Miao,G., Caraher,N. and Hanafey,M.K.
  TITLE     DuPont Wheat cDNA Sequence
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Scott V. Tingey
            Crop Genetics
            E. I. DuPont de Nemours and Company
            1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
            Tel: 302-631-2602
            Fax: 302-631-2607
            Email: Scott.V.Tingey@USA.dupont.com
            Seq primer: M13.
FEATURES              Location/Qualifiers
     source           1. .562
                      /organism="Triticum aestivum"
                      /mol_type="mRNA"
                      /db_xref="taxon:4565"
                      /clone="wlp1c.pk006.d15"
                      /tissue_type="lemma and palea"
                      /lab_host="DH10B"
                      /clone_lib="wlp1c"
                      /note="Vector: pBluescript SK+; Site_1: EcoRI; Site_2:
                      XhoI; Wheat (Triticum aestivum, Hi Line) lemma and palea"
ORIGIN

  Query Match             0.5%;  Score 45.6;  DB 6;  Length 562;
  Best Local Similarity   55.8%;  Pred. No. 1.8;
  Matches   87;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy       7167 CTCTTCCATGCCCCCCCCTCGAGGGGGAGCTTGGAGATCCAGACCTGGAGCCTGAGCAGGT 7226
              |||   ||||| |  |||||||| |||| || | |      |||||||  ||   ||||
Db        318 CTCCGTCATGCGCAGCCTCGAGGAGGAGATTTGCGCCGGCGACCTGGCCCCGCCGCAGCC 377

Qy       7227 AGAGCCCCAACCCCCCCCCCAGGGGGGGGTGGCAGCTCCCGGCTCGGACTCGGGGTCCTG 7286
              |||| |     ||  | | |||    ||  |  |  | ||||  |    ||| | |  |
Db        378 GGAGCTCGGCTTCCTGCTCGAGGCCTCGGACGACGAGCTCGGCCTGCCGCCGGCGGCGGG 437

Qy       7287 GTCTACTTGCTCCGAGGAGGACGACTCCGTCGTGTG 7322
                |  | | ||||  |  |||||||| |||  |  ||
Db        438 CGCGGCGTCCTCCTCGTCGGACGACGCCGGGGGCTG 473


Search completed: November  1, 2005, 04:05:44
Job time : 17327 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 31, 2005, 15:25:30 ; Search time 937 Seconds
(without alignments)
15889.522 Million cell updates/sec

Title: US-09-980-559-1_COPY_341_9439
Perfect score: 9099
Sequence: 1 atgagcacaaatcctaaacc.....ttttcctactccccgctcgg 9099

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

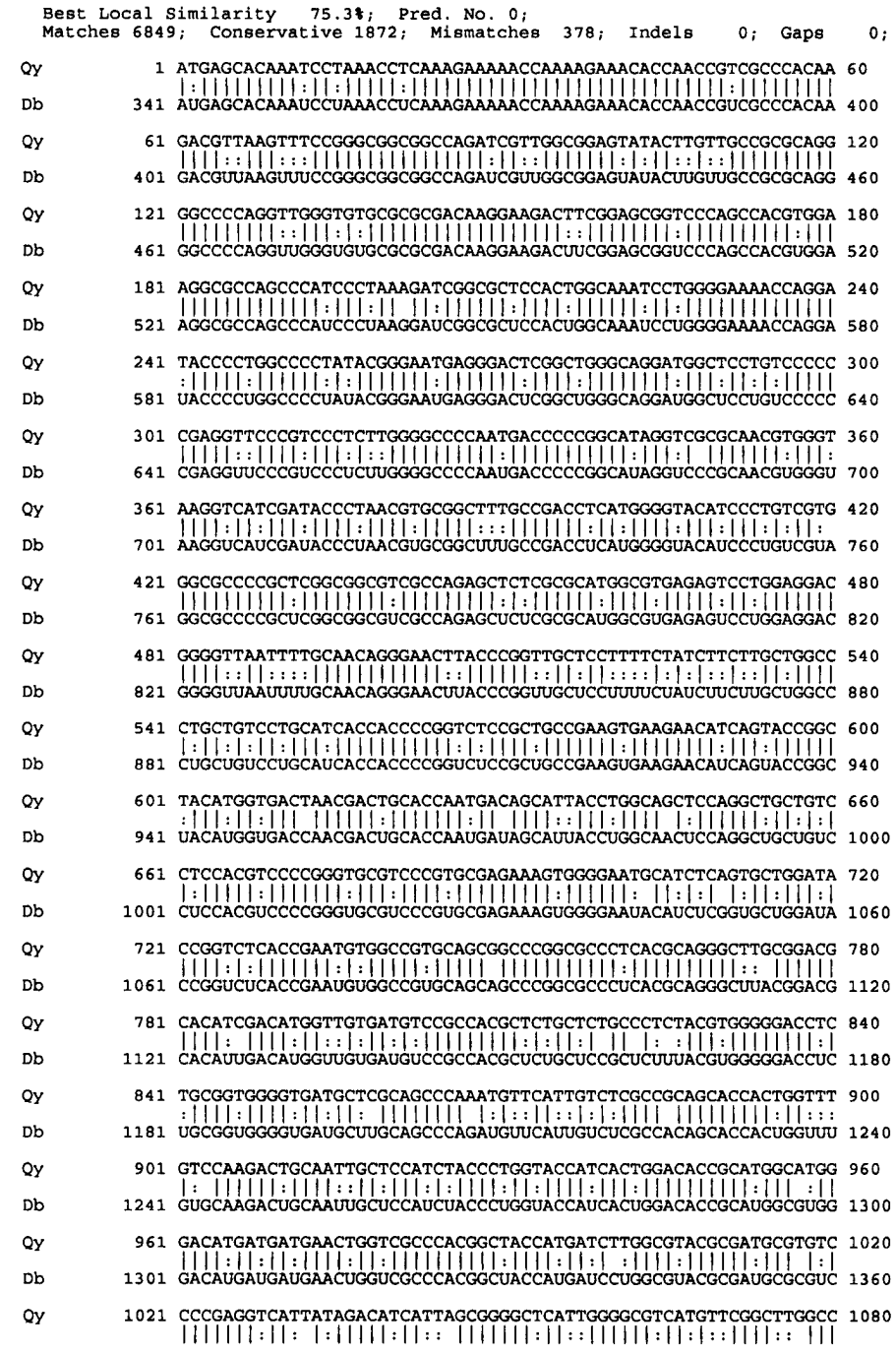
Result No.	Score	% Query Match	Length	DB	ID	Description
1	8494.2	93.4	9589	1	US-07-925-695-1	Sequence 1, Appli
2	8494.2	93.4	9589	1	US-07-925-695-2	Sequence 2, Appli
3	5653.8	62.1	9511	1	US-07-925-695-7	Sequence 7, Appli
4	5651	62.1	9511	1	US-07-925-695-6	Sequence 6, Appli
5	4141.8	45.5	9413	4	US-09-827-688-6	Sequence 6, Appli
6	4136	45.5	9599	3	US-09-014-416-6	Sequence 6, Appli
7	4129.6	45.4	12980	3	US-08-811-566-5	Sequence 5, Appli
8	4129.6	45.4	12980	3	US-09-034-756-5	Sequence 5, Appli
9	4128	45.4	9646	3	US-08-811-566-1	Sequence 1, Appli
10	4128	45.4	9646	3	US-09-034-756-1	Sequence 1, Appli
11	4121.6	45.3	9599	3	US-09-014-416-2	Sequence 2, Appli
12	4118	45.3	9401	5	PCT-US91-02225-9	Sequence 9, Appli
13	4115.2	45.2	9379	3	US-09-388-874-1	Sequence 1, Appli
14	4115.2	45.2	9379	4	US-09-916-359-1	Sequence 1, Appli
15	4115.2	45.2	9401	1	US-07-910-760-9	Sequence 9, Appli
16	4115.2	45.2	9401	1	US-08-440-519-9	Sequence 9, Appli
17	4115.2	45.2	9401	3	US-08-440-549-9	Sequence 9, Appli
18	4115.2	45.2	9401	3	US-08-823-895A-25	Sequence 25, Appl
19	4114.8	45.2	9379	3	US-08-444-818-176	Sequence 176, App
20	4112	45.2	9365	4	US-09-827-688-7	Sequence 7, Appli
21	4106	45.1	9030	1	US-08-324-977-13	Sequence 13, Appl
22	4106	45.1	9030	2	US-08-384-616-13	Sequence 13, Appl
23	4106	45.1	9030	2	US-08-904-686A-13	Sequence 13, Appl
24	4106	45.1	9030	3	US-09-315-850-13	Sequence 13, Appl
25	4106	45.1	9416	1	US-08-324-977-1	Sequence 1, Appli
26	4106	45.1	9416	2	US-08-384-616-1	Sequence 1, Appli
27	4106	45.1	9416	2	US-08-904-686A-1	Sequence 1, Appli
28	4106	45.1	9416	3	US-09-315-850-1	Sequence 1, Appli
29	4106	45.1	9416	3	US-08-823-895A-27	Sequence 27, Appl
30	4103.4	45.1	9401	2	US-08-432-693-1	Sequence 1, Appli
31	4103.4	45.1	9416	3	US-08-811-566-19	Sequence 19, Appl
32	4103.4	45.1	9416	3	US-09-034-756-19	Sequence 19, Appl
33	4101.4	45.1	9595	3	US-09-014-416-4	Sequence 4, Appli
34	4093.2	45.0	11076	4	US-09-539-601-1	Sequence 1, Appli
35	4083.6	44.9	8987	3	US-08-444-818-137	Sequence 137, App
36	4083.6	44.9	11076	4	US-09-539-601-25	Sequence 25, Appl
37	4082	44.9	11076	4	US-09-539-601-19	Sequence 19, Appl
38	4080	44.8	9416	3	US-08-823-895A-26	Sequence 26, Appl
39	4080	44.8	9416	4	US-10-104-966-13	Sequence 13, Appl
40	4077.2	44.8	11076	4	US-09-539-601-31	Sequence 31, Appl
41	4048.4	44.5	9472	4	US-08-150-204E-96	Sequence 96, Appl
42	4014.8	44.1	9185	3	US-08-444-818-122	Sequence 122, App
c 43	4014.8	44.1	9185	3	US-08-444-818-123	Sequence 123, App
44	3723.6	40.9	3970	1	US-07-925-695-3	Sequence 3, Appli
45	3678.4	40.4	8316	3	US-08-444-818-88	Sequence 88, Appl

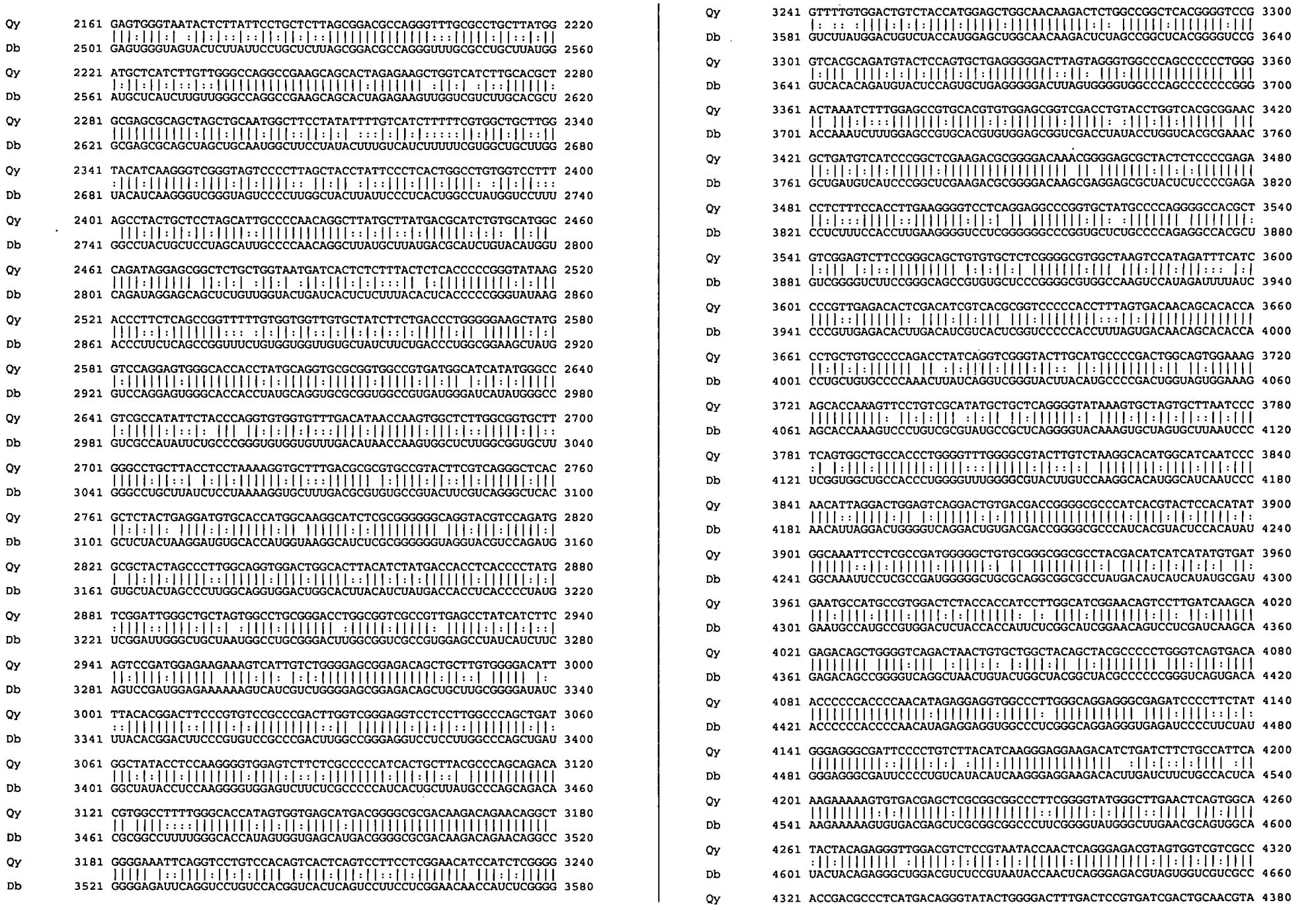
ALIGNMENTS

RESULT 1
US-07-925-695-1
; Sequence 1, Application US/07925695
; Patent No. 5428145
; GENERAL INFORMATION:
; APPLICANT: OKAMOTO, Hiroaki
; APPLICANT: NAKAMURA, Tetsuo
; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOME,
; TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, ANTIGEN, ANTIBODY AND
; TITLE OF INVENTION: DETECTION SYSTEMS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Beveridge, DeGrandi, Weilacher & Young
; STREET: 1850 M Street, N.W., Suite 800
; CITY: Washington
; STATE: D.C.
; COUNTRY: US
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/925,695
; FILING DATE: 19920807
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 287402/91
; FILING DATE: 09-AUG-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 360441/91
; FILING DATE: 05-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Weilacher, Robert G.
; REGISTRATION NUMBER: 20,531
; REFERENCE/DOCKET NUMBER: 06/87-48009
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 659-2811
; TELEFAX: (202) 659-1462
; TELEX: WUI 64470
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9589 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
US-07-925-695-1

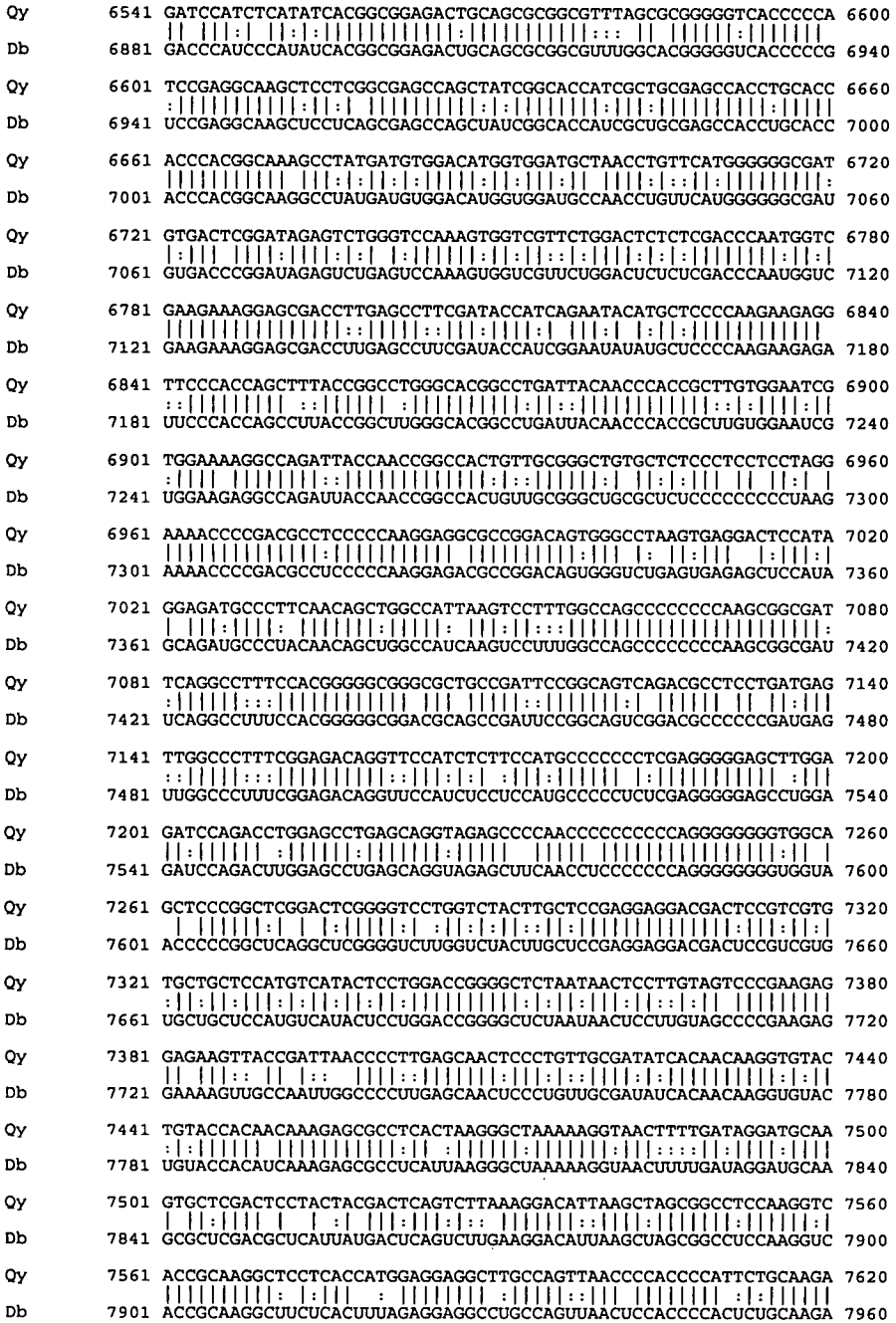
Query Match 93.4%; Score 8494.2; DB 1; Length 9589;

Best Local Similarity 75.3%; Pred. No. 0;
Matches 6849; Conservative 1872; Mismatches 378; Indels 0; Gaps 0;








Qy 6541 GATCCATCTCATATCACGGCGGAGACTGCAGCGCGGCGTTTAGCGCGGGGGTCACCCCCA 6600
Db 6881 GACCCAUCCCAUAUCACGGCGGAGACUGCAGCGCGGCGUUUGGCACGGGGGUCACCCCCG 6940
Qy 6601 TCCGAGGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 6660
Db 6941 UCCGAGGCAAGCUCCUCAGCGAGCCAGCUAUCGGCACCAUCGCUGCGAGCCACCUGCACC 7000
Qy 6661 ACCCACGGCAAAGCCTATGATGTGGACATGGTGGATGCTAACCTGTTCATGGGGGGCGAT 6720
Db 7001 ACCCACGGCAAGGCCUAUGAUGUGGACAUGGUGGAUGCCAACCUGUUCAUGGGGGGCGAU 7060
Qy 6721 GTGACTCGGATAGAGTCTGGGTCCAAAGTGGTCGTTCTGGACTCTCTCGACCCAATGGTC 6780
Db 7061 GUGACCCGGAUAGAGUCUGAGUCCAAAGUGGUCGUUCUGGACUCUCUCGACCCAAUGGUC 7120
Qy 6781 GAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCAGAATACATGCTCCCCAAGAAGAGG 6840
Db 7121 GAAGAAAGGAGCGACCUUGAGCCUUCGAUACCAUCGGAAUAUAUGCUCCCCAAGAAGAGA 7180
Qy 6841 TTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCGCTTGTGGAATCG 6900
Db 7181 UUCCCACCAGCCUUACCGGCUUGGGCACGGCCUGAUUACAACCCACCGCUUGUGGAAUCG 7240
Qy 6901 TGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTCCCTCCTCCTAGG 6960
Db 7241 UGGAAGAGGCCAGAUUACCAACCGGCCACUGUUGCGGGCUGCGCUCUCCCCCCCCCUAAG 7300
Qy 6961 AAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGTGAGGACTCCATA 7020
Db 7301 AAAACCCCGACGCCUCCCCCAAGGAGACGCCGGACAGUGGGUCUGAGUGAGAGCUCCAUA 7360
Qy 7021 GGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCCCCCAAGCGGCGAT 7080
Db 7361 GCAGAUGCCCUACAACAGCUGGCCAUCAAGUCCUUUGGCCAGCCCCCCCCCAAGCGGCGAU 7420
Qy 7081 TCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACGCCTCCTGATGAG 7140
Db 7421 UCAGGCCUUUCCACGGGGGCGGACGCAGCCGAUUCCGGCAGUCGGACGCCCCCCGAUGAG 7480
Qy 7141 TTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCTCGAGGGGGAGCTTGGA 7200
Db 7481 UUGGCCCUUUCGGAGACAGGUUCCAUCUCCUCCAUGCCCCCUCUCGAGGGGGAGCCUGGA 7540
Qy 7201 GATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCCAGGGGGGGGTGGCA 7260
Db 7541 GAUCCAGACUUGGAGCCUGAGCAGGUAGAGCUUCAACCUCCCCCCCAGGGGGGGGUGGUA 7600
Qy 7261 GCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGACGACTCCGTCGTG 7320
Db 7601 ACCCCCGGCUCAGGCUCGGGGUCUUGGUCUACUUGCUCCGAGGAGGACGACUCCGUCGUG 7660
Qy 7321 TGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGTAGTCCCGAAGAG 7380
Db 7661 UGCUGCUCCAUGUCAUACUCCUGGACCGGGGCUCUAAUAACUCCUUGUAGCCCCGAAGAG 7720
Qy 7381 GAGAAGTTACCGATTAACCCCTTGAGCAACTCCCTGTTGCGATATCACAACAAGGTGTAC 7440
Db 7721 GAAAAGUUGCCAAUUGGCCCCUUGAGCAACUCCCUGUUGCGAUAUCACAACAAGGUGUAC 7780
Qy 7441 TGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTTGATAGGATGCAA 7500
Db 7781 UGUACCACAUCAAAGAGCGCCUCAUUAAGGGCUAAAAAGGUAACUUUUGAUAGGAUGCAA 7840
Qy 7501 GTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCGGCCTCCAAGGTC 7560
Db 7841 GCGCUCGACGCUCAUUAUGACUCAGUCUUGAAGGACAUUAAGCUAGCGGCCUCCAAGGUC 7900
Qy 7561 ACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCCCATTCTGCAAGA 7620
Db 7901 ACCGCAAGGCUUCUCACUUUAGAGGAGGCCUGCCAGUUAACUCCACCCCACUCUGCAAGA 7960
Qy 7621 TCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGGGCCGTTAACCAC 7680
Db 7961 UCCAAGUAUGGGUUUGGGGCUAAGGAGGUCCGCAGCUUGUCCGGGAGAGCCGUUAACCAC 8020
Qy 7681 ATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATTCCCACAACCATT 7740
Db 8021 AUCAAGUCCGUGUGGAAGGACCUCCUGGAAGACACACAAACACCAAUUCCUACAACCAUC 8080
Qy 7741 ATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAGAAAGCAGCTCGC 7800
Db 8081 AUGGCCAAAAAUGAGGUGUUCUGCGUGGACCCCACCAAGGGGGGUAAGAAAGCAGCUCGC 8140
Qy 7801 CTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCCCTTTATGACATT 7860
Db 8141 CUUAUCGUUUACCCUGACCUCGGCGUCAGGGUCUGCGAGAAAAUGGCCCUUUAUGAUAUC 8200
Qy 7861 ACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAGTATTCCCCCGCT 7920
Db 8201 ACACAAAAGCUUCCUCAGGCGGUGAUGGGGGCUUCUUAUGGAUUCCAGUACUCCCCCGCU 8260
Qy 7921 CAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCTATGGGTTTTTCG 7980
Db 8261 CAGCGGGUGGAGUUUCUCUUGAAGGCAUGGGCGGAAAAGAAAGACCCUAUGGGUUUUUCG 8320
Qy 7981 TATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGGACTGAGGAGTCC 8040
Db 8321 UAUGAUACCCGAUGCUUUGACUCAACCGUCACUGAGAGAGACAUCAGGACUGAGGAGUCC 8380
Qy 8041 ATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACACTCGCTAACTGAG 8100
Db 8381 AUAUAUCGGGCUUGUUCCUUGCCCGAGGAGGCCCACACUGCCAUACACUCACUGACUGAG 8440
Qy 8101 AGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGCGGGTACAGGCGT 8160
Db 8441 AGACUUUACGUGGGAGGGCCCAUGUUCAACAGCAAGGGCCAGACCUGCGGGUACAGGCGU 8500
Qy 8161 TGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACATGCTACGTGAAA 8220
Db 8501 UGCCGCGCCAGCGGGGUGCUUACCACUAGCAUGGGGAACACCAUCACAUGCUAUGUGAAA 8560
Qy 8221 GCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTGGTATGCGGCGAT 8280
Db 8561 GCCUUAGCGGCCUGUAAGGCUGCAGGGAUAAUUGCGCCCACAAUGCUGGUAUGCGGCGAU 8620
Qy 8281 GACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGGAACCTGAGAGCC 8340
Db 8621 GACUUGGUUGUCAUCUCAGAGAGCCAGGGGACCGAGGAGGACGAGCGGAACCUGAGAGCC 8680
Qy 8341 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCCAGACCGGAGTAT 8400
Db 8681 UUCACGGAGGCUAUGACCAGGUAUUCUGCCCCUCCUGGUGACCCCCCCAGACCGGAAUAU 8740
Qy 8401 GATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTGGGCCCACAAGGC 8460
Db 8741 GACCUGGAGCUGAUAACAUCUUGCUCCUCAAAUGUGUCUGUGGCGUUGGGCCCACAAGGC 8800
Qy 8461 CGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGGGCTGCCTGGGAA 8520
Db 8801 CGCCGCAGAUACUACCUGACCAGAGACCCUACCACUCCAAUCGCCCGGGCUGCCUGGGAA 8860
Qy 8521 ACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAGTACGCCCCGACC 8580
Db 8861 ACAGUUAGACACUCCCCUGUCAAUUCAUGGCUAGGAAACAUCAUCCAGUACGCCCCAACC 8920
Qy 8581 ATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATGGCTCAAGACACG 8640
Db 8921 AUAUGGGCUCGCAUGGUCCUGAUGACACACUUCUUCUCCAUUCUCAUGGCCCAAGAUACU 8980
Qy 8641 CTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTGAGTCCCTTGGAC 8700
Db 8981 CUGGACCAGAACCUCAACUUUGAGAUGUACGGAGCGGUGUACUCCGUGAGUCCCUUGGAC 9040
Qy 8701 CTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTGCACACATACACT 8760

```
Db       9041 CUCCCAGCCAUAAUUGAAAGGUUACACGGGCUUGACGCUUUCUCUCUGCACACAUACACU 9100

Qy       8761 CCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGA 8820

Db       9101 CCCCACGAACUGACACGGGUGGCUUCAGCCCUCAGAAAACUUGGGGCGCCACCCCUCAGA 9160

Qy       8821 GCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGTGGGGGGAGAGCG 8880

Db       9161 GCGUGGAAGAGCCGGGCACGUGCAGUCAGGGCGUCCCUCAUCUCCCGUGGGGGGAGAGCG 9220

Qy       8881 GCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 8940

Db       9221 GCCGUUUGCGGCCGAUAUCUCUUCAACUGGGCGGUGAAGACCAAGCUCAAACUCACUCCA 9280

Qy       8941 TTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGCGCCGGCGGGGGC 9000

Db       9281 UUGCCGGAAGCGCGCCUCCUGGAUUUAUCCAGCUGGUUCACUGUCGGCGCCGGCGGGGGC 9340

Qy       9001 GACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTTGGCCTACTCCTA 9060

Db       9341 GACAUUUAUCACAGCGUGUCGCGUGCCCGACCCCGCUUAUUACUCCUUGGCCUACUCCUA 9400

Qy       9061 CTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCGG 9099

Db       9401 CUUUUUGUAGGGGUAGGCCUUUUCCUACUCCCCGCUCGG 9439


RESULT 2
US-07-925-695-2
; Sequence 2, Application US/07925695
; Patent No. 5428145
;  GENERAL INFORMATION:
;    APPLICANT:  OKAMOTO, Hiroaki
;    APPLICANT:  NAKAMURA, Tetsuo
;    TITLE OF INVENTION:  NON-A, NON-B HEPATITIS VIRUS GENOME,
;    TITLE OF INVENTION:  POLYNUCLEOTIDES, POLYPEPTIDES, ANTIGEN, ANTIBODY AND
;    TITLE OF INVENTION:  DETECTION SYSTEMS
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Beveridge, DeGrandi, Weilacher & Young
;      STREET:  1850 M Street, N.W., Suite 800
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  US
;      ZIP:  20036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/925,695
;      FILING DATE:  19920807
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 287402/91
;      FILING DATE:  09-AUG-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 360441/91
;      FILING DATE:  05-DEC-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Weilacher, Robert G.
;      REGISTRATION NUMBER:  20,531
;      REFERENCE/DOCKET NUMBER:  06/87-48009
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 659-2811
;      TELEFAX:  (202) 659-1462
;      TELEX:  WUI 64470
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9589 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-07-925-695-2

  Query Match             93.4%;  Score 8494.2;  DB 1;  Length 9589;
  Best Local Similarity   95.8%;  Pred. No. 0;
  Matches 8721;  Conservative    0;  Mismatches  378;  Indels    0;  Gaps    0;

Qy          1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 60

Db        341 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 400

Qy         61 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 120

Db        401 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 460

Qy        121 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 180

Db        461 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 520

Qy        181 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 240

Db        521 AGGCGCCAGCCCATCCCTAAGGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 580

Qy        241 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 300

Db        581 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 640

Qy        301 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 360

Db        641 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCCCGCAACGTGGGT 700

Qy        361 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 420

Db        701 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTA 760

Qy        421 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 480

Db        761 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 820

Qy        481 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 540

Db        821 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 880

Qy        541 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 600

Db        881 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 940

Qy        601 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 660

Db        941 TACATGGTGACCAACGACTGCACCAATGATAGCATTACCTGGCAACTCCAGGCTGCTGTC 1000

Qy        661 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 720

Db       1001 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATACATCTCGGTGCTGGATA 1060

Qy        721 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 780

Db       1061 CCGGTCTCACCGAATGTGGCCGTGCAGCAGCCCGGCGCCCTCACGCAGGGCTTACGGACG 1120

Qy        781 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 840

Db       1121 CACATTGACATGGTTGTGATGTCCGCCACGCTCTGCTCCGCTCTTTACGTGGGGGACCTC 1180

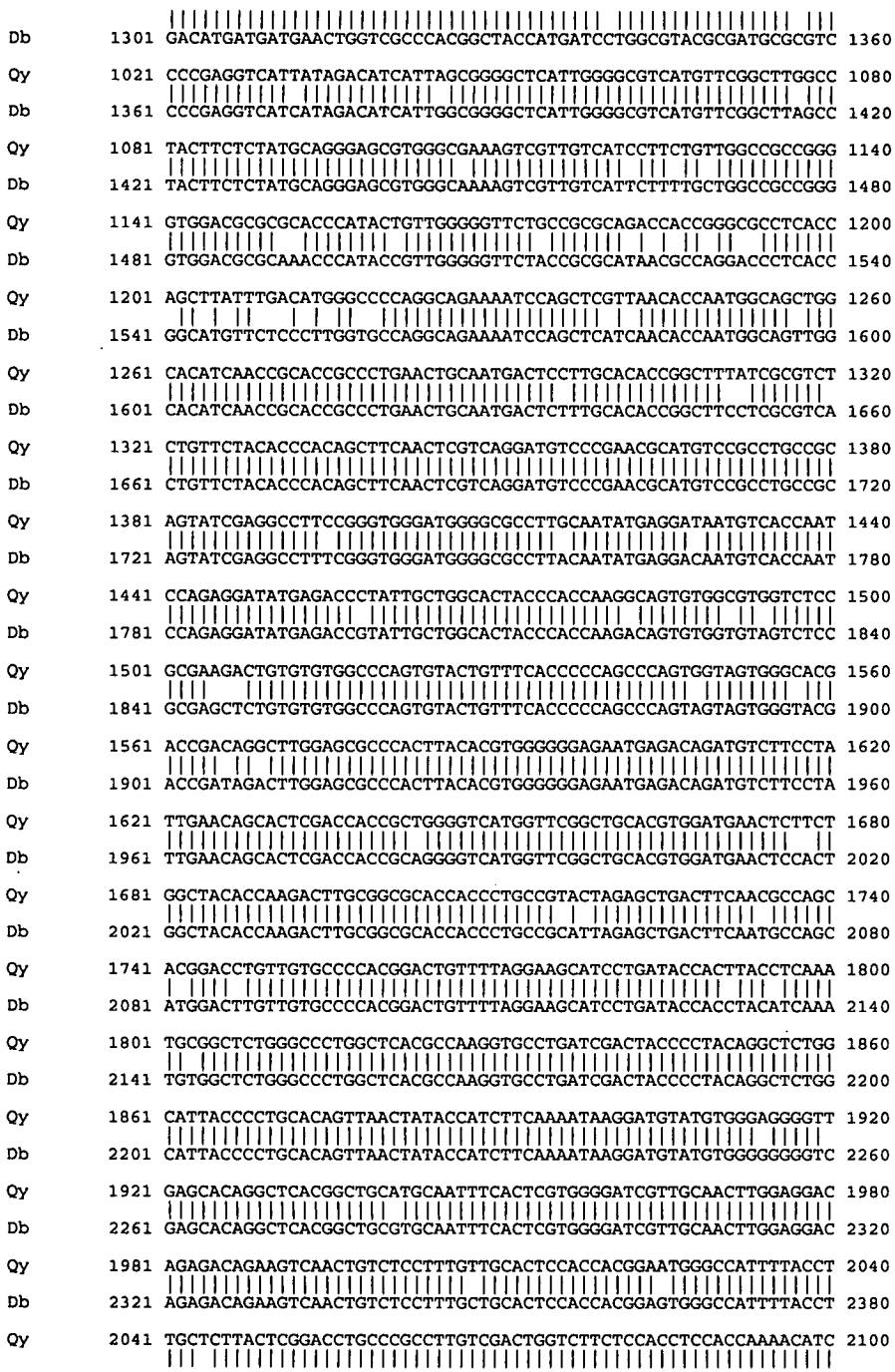
Qy        841 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 900

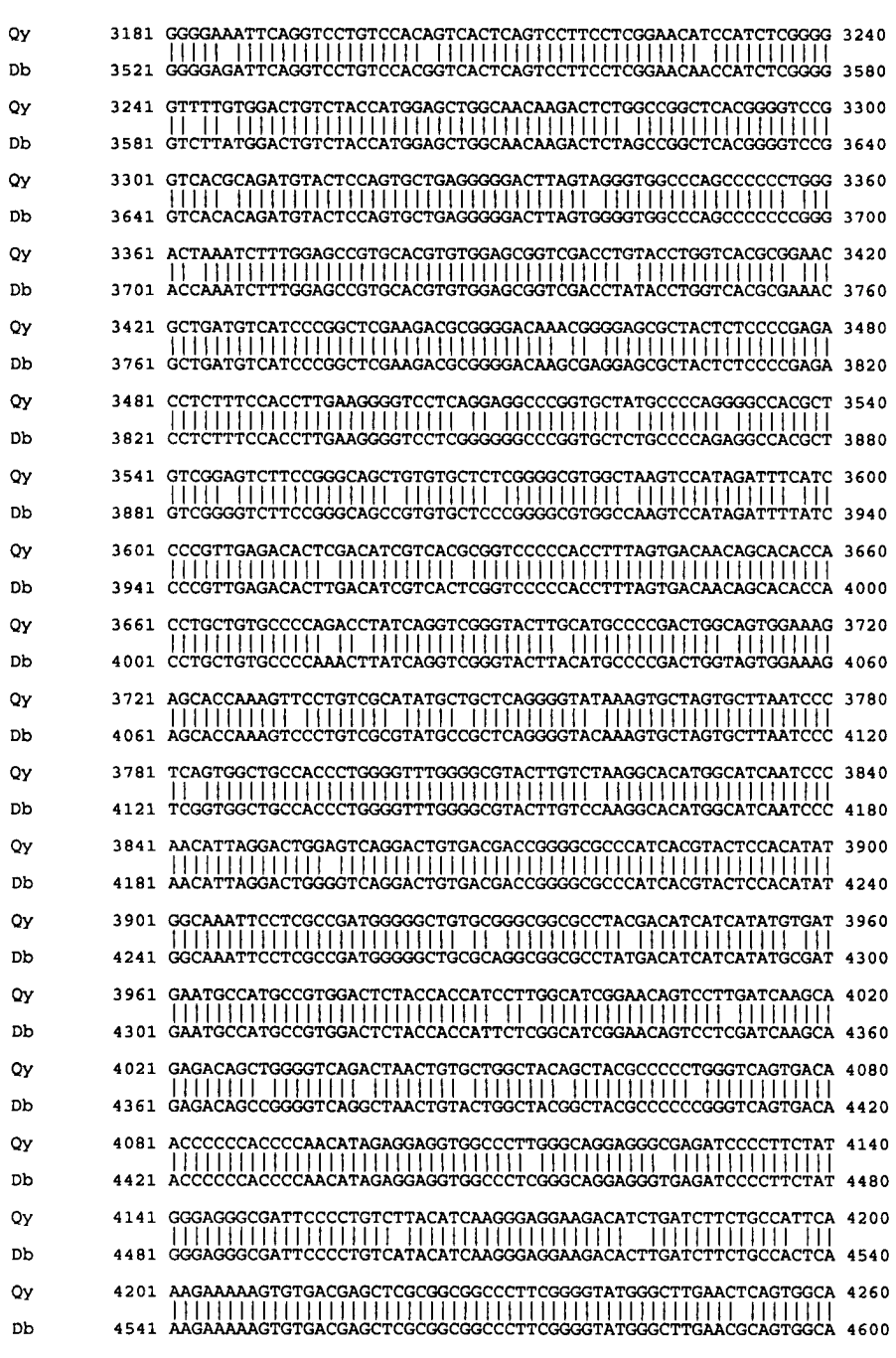
Db       1181 TGCGGTGGGGTGATGCTTGCAGCCCAGATGTTCATTGTCTCGCCACAGCACCACTGGTTT 1240

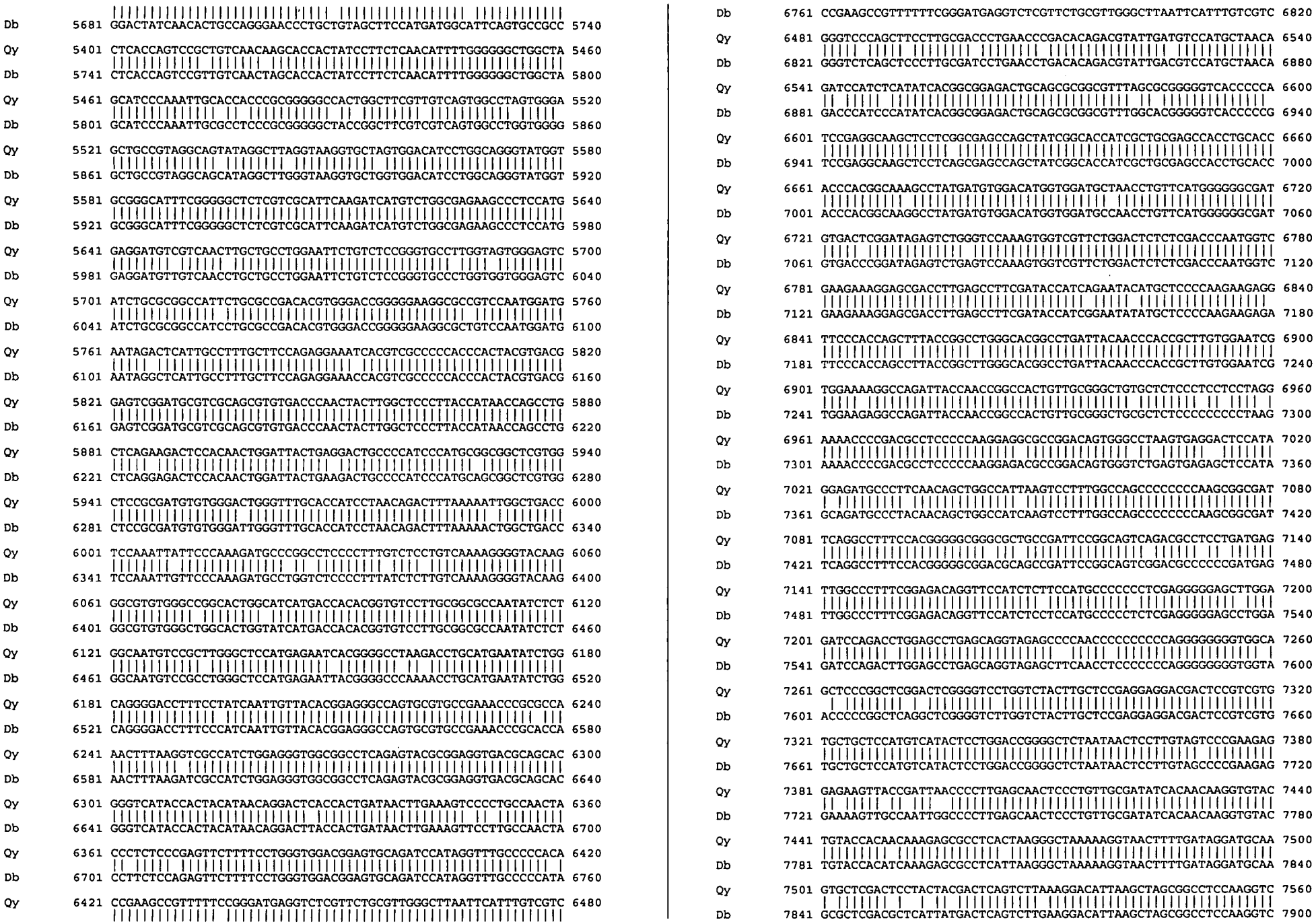
Qy        901 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 960

Db       1241 GTGCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCGTGG 1300

Qy        961 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1020
```


Db 1301 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCCTGGCGTACGCGATGCGCGTC 1360
Qy 1021 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1080
Db 1361 CCCGAGGTCATCATAGACATCATTGGCGGGGCTCATTGGGGCGTCATGTTCGGCTTAGCC 1420
Qy 1081 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1140
Db 1421 TACTTCTCTATGCAGGGAGCGTGGGCAAAAGTCGTTGTCATTCTTTTGCTGGCCGCCGGG 1480
Qy 1141 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1200
Db 1481 GTGGACGCGCAAACCCATACCGTTGGGGGTTCTACCGCGCATAACGCCAGGACCCTCACC 1540
Qy 1201 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1260
Db 1541 GGCATGTTCTCCCTTGGTGCCAGGCAGAAAATCCAGCTCATCAACACCAATGGCAGTTGG 1600
Qy 1261 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1320
Db 1601 CACATCAACCGCACCGCCCTGAACTGCAATGACTCTTTGCACACCGGCTTCCTCGCGTCA 1660
Qy 1321 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1380
Db 1661 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1720
Qy 1381 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1440
Db 1721 AGTATCGAGGCCTTTCGGGTGGGATGGGGCGCCTTACAATATGAGGACAATGTCACCAAT 1780
Qy 1441 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1500
Db 1781 CCAGAGGATATGAGACCGTATTGCTGGCACTACCCACCAAGACAGTGTGGTGTAGTCTCC 1840
Qy 1501 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1560
Db 1841 GCGAGCTCTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTAGTAGTGGGTACG 1900
Qy 1561 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1620
Db 1901 ACCGATAGACTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1960
Qy 1621 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 1680
Db 1961 TTGAACAGCACTCGACCACCGCAGGGGTCATGGTTCGGCTGCACGTGGATGAACTCCACT 2020
Qy 1681 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 1740
Db 2021 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGCATTAGAGCTGACTTCAATGCCAGC 2080
Qy 1741 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 1800
Db 2081 ATGGACTTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACCTACATCAAA 2140
Qy 1801 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 1860
Db 2141 TGTGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 2200
Qy 1861 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 1920
Db 2201 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGGGGGGTC 2260
Qy 1921 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 1980
Db 2261 GAGCACAGGCTCACGGCTGCGTGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 2320
Qy 1981 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2040
Db 2321 AGAGACAGAAGTCAACTGTCTCCTTTGCTGCACTCCACCACGGAGTGGGCCATTTTACCT 2380
Qy 2041 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2100
Db 2381 TGCACTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2440
Qy 2101 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2160
Db 2441 GTGGACGTGCAATTCATGTATGGCCTATCACCTGCTCTCACAAAATACATCGTCCGATGG 2500
Qy 2161 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2220
Db 2501 GAGTGGGTAGTACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2560
Qy 2221 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2280
Db 2561 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGTTGGTCGTCTTGCACGCT 2620
Qy 2281 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATATTTTGTCATCTTTTTCGTGGCTGCTTGG 2340
Db 2621 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATACTTTGTCATCTTTTTCGTGGCTGCTTGG 2680
Qy 2341 TACATCAAGGGTCGGGTAGTCCCCTTAGCTACCTATTCCCTCACTGGCCTGTGGTCCTTT 2400
Db 2681 TACATCAAGGGTCGGGTAGTCCCCTTGGCTACTTATTCCCTCACTGGCCTATGGTCCTTT 2740
Qy 2401 AGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTGCATGGC 2460
Db 2741 GGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTACATGGT 2800
Qy 2461 CAGATAGGAGCGGCTCTGCTGGTAATGATCACTCTCTTTACTCTCACCCCCGGGTATAAG 2520
Db 2801 CAGATAGGAGCAGCTCTGTTGGTACTGATCACTCTCTTTACACTCACCCCCGGGTATAAG 2860
Qy 2521 ACCCTTCTCAGCCGGTTTTTGTGGTGGTTGTGCTATCTTCTGACCCTGGGGGAAGCTATG 2580
Db 2861 ACCCTTCTCAGCCGGTTTCTGTGGTGGTTGTGCTATCTTCTGACCCTGGCGGAAGCTATG 2920
Qy 2581 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGCATCATATGGGCC 2640
Db 2921 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGGATCATATGGGCC 2980
Qy 2641 GTCGCCATATTCTACCCAGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 2700
Db 2981 GTCGCCATATTCTGCCCGGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 3040
Qy 2701 GGGCCTGCTTACCTCCTAAAAGGTGCTTTGACGCGCGTGCCGTACTTCGTCAGGGCTCAC 2760
Db 3041 GGGCCTGCTTATCTCCTAAAAGGTGCTTTGACGCGTGTGCCGTACTTCGTCAGGGCTCAC 3100
Qy 2761 GCTCTACTGAGGATGTGCACCATGGCAAGGCATCTCGCGGGGGGCAGGTACGTCCAGATG 2820
Db 3101 GCTCTACTAAGGATGTGCACCATGGTAAGGCATCTCGCGGGGGGTAGGTACGTCCAGATG 3160
Qy 2821 GCGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 2880
Db 3161 GTGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 3220
Qy 2881 TCGGATTGGGCTGCTAGTGGCCTGCGGGACCTGGCGGTCGCCGTTGAGCCTATCATCTTC 2940
Db 3221 TCGGATTGGGCTGCTAATGGCCTGCGGGACTTGGCGGTCGCCGTGGAGCCTATCATCTTC 3280
Qy 2941 AGTCCGATGGAGAAGAAAGTCATTGTCTGGGGAGCGGAGACAGCTGCTTGTGGGGACATT 3000
Db 3281 AGTCCGATGGAGAAAAAAGTCATCGTCTGGGGAGCGGAGACAGCTGCTTGCGGGGATATC 3340
Qy 3001 TTACACGGACTTCCCGTGTCCGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3060
Db 3341 TTACACGGACTTCCCGTGTCCGCCCGACTTGGCCGGGAGGTCCTCCTTGGCCCAGCTGAT 3400
Qy 3061 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3120
Db 3401 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTATGCCCAGCAGACA 3460
Qy 3121 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCT 3180
Db 3461 CGCGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCC 3520


Qy 3181 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3240
Db 3521 GGGGAGATTCAGGTCCTGTCCACGGTCACTCAGTCCTTCCTCGGAACAACCATCTCGGGG 3580
Qy 3241 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3300
Db 3581 GTCTTATGGACTGTCTACCATGGAGCTGGCAACAAGACTCTAGCCGGCTCACGGGGTCCG 3640
Qy 3301 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGG 3360
Db 3641 GTCACACAGATGTACTCCAGTGCTGAGGGGGACTTAGTGGGGTGGCCCAGCCCCCCCGGG 3700
Qy 3361 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3420
Db 3701 ACCAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTATACCTGGTCACGCGAAAC 3760
Qy 3421 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3480
Db 3761 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAGCGAGGAGCGCTACTCTCCCCGAGA 3820
Qy 3481 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3540
Db 3821 CCTCTTTCCACCTTGAAGGGGTCCTCGGGGGGCCCGGTGCTCTGCCCCAGAGGCCACGCT 3880
Qy 3541 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3600
Db 3881 GTCGGGGTCTTCCGGGCAGCCGTGTGCTCCCGGGGCGTGGCCAAGTCCATAGATTTTATC 3940
Qy 3601 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 3660
Db 3941 CCCGTTGAGACACTTGACATCGTCACTCGGTCCCCCACCTTTAGTGACAACAGCACACCA 4000
Qy 3661 CCTGCTGTGCCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 3720
Db 4001 CCTGCTGTGCCCCAAACTTATCAGGTCGGGTACTTACATGCCCCGACTGGTAGTGGAAAG 4060
Qy 3721 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 3780
Db 4061 AGCACCAAAGTCCCTGTCGCGTATGCCGCTCAGGGGTACAAAGTGCTAGTGCTTAATCCC 4120
Qy 3781 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 3840
Db 4121 TCGGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCCAAGGCACATGGCATCAATCCC 4180
Qy 3841 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 3900
Db 4181 AACATTAGGACTGGGGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 4240
Qy 3901 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 3960
Db 4241 GGCAAATTCCTCGCCGATGGGGGCTGCGCAGGCGGCGCCTATGACATCATCATATGCGAT 4300
Qy 3961 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4020
Db 4301 GAATGCCATGCCGTGGACTCTACCACCATTCTCGGCATCGGAACAGTCCTCGATCAAGCA 4360
Qy 4021 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4080
Db 4361 GAGACAGCCGGGGTCAGGCTAACTGTACTGGCTACGGCTACGCCCCCCGGGTCAGTGACA 4420
Qy 4081 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4140
Db 4421 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTCGGGCAGGAGGGTGAGATCCCCTTCTAT 4480
Qy 4141 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4200
Db 4481 GGGAGGGCGATTCCCCTGTCATACATCAAGGGAGGAAGACACTTGATCTTCTGCCACTCA 4540
Qy 4201 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4260
Db 4541 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACGCAGTGGCA 4600
Qy 4261 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4320
Db 4601 TACTACAGAGGGCTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4660
Qy 4321 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4380
Db 4661 ACCGACGCCCTCATGACGGGGTTTACTGGAGACTTTGACTCCGTGATCGACTGCAACGTA 4720
Qy 4381 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4440
Db 4721 GCGGTCACTCAAGTTGTAGACTTCAGCTTGGACCCCACATTCACCATAACCACACAGACT 4780
Qy 4441 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4500
Db 4781 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGCCGCACGGGCAGGGGAAGACTG 4840
Qy 4501 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4560
Db 4841 GGTATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4900
Qy 4561 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4620
Db 4901 CTCTGCGAGTGCTACGATGCAGGGGCCGCATGGTATGAGCTCACACCAGCGGAGACCACC 4960
Qy 4621 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 4680
Db 4961 GTCAGGCTCAGAGCATATTTCAACACACCTGGTTTGCCTGTGTGCCAAGACCATCTTGAG 5020
Qy 4681 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 4740
Db 5021 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 5080
Qy 4741 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 4800
Db 5081 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 5140
Qy 4801 GCCAAAGCCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 4860
Db 5141 GCCAAAGCCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 5200
Qy 4861 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 4920
Db 5201 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 5260
Qy 4921 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 4980
Db 5261 CTCACGCATCCTGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 5320
Qy 4981 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5040
Db 5321 ACCAGCACGTGGGTCTTAGCTGGGGGGGTCTTGGCGGCCGTCGCCGCGTACTGCCTGGCG 5380
Qy 5041 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5100
Db 5381 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACGTTAACCAGCGAGCCGTCGTTGCACCG 5440
Qy 5101 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5160
Db 5441 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGAGCGGCT 5500
Qy 5161 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5220
Db 5501 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5560
Qy 5221 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5280
Db 5561 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCGCTGTGCAGGCTTCTTGGCCCAAG 5620
Qy 5281 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5340
Db 5621 GTAGAGCAATTCTGGGCCAAACACATGTGGAACTTCATCAGCGGCATTCAATACCTCGCA 5680
Qy 5341 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5400



```
Qy       7561 ACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCCCATTCTGCAAGA 7620
             |||||||||||| |||||  | |||||||| |||||||||| |||||||| |||||||||
Db       7901 ACCGCAAGGCTTCTCACTTTAGAGGAGGCCTGCCAGTTAACTCCACCCCACTCTGCAAGA 7960

Qy       7621 TCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGGGCCGTTAACCAC 7680
             || || ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db       7961 TCCAAGTATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGAGCCGTTAACCAC 8020

Qy       7681 ATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATTCCCACAACCATT 7740
             |||||||||||||||||||||||||||| ||| || |||||||||||| ||||||||
Db       8021 ATCAAGTCCGTGTGGAAGGACCTCCTGGAAGACACACAAACACCAATTCCTACAACCATC 8080

Qy       7741 ATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAGAAAGCAGCTCGC 7800
             |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db       8081 ATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGTAAGAAAGCAGCTCGC 8140

Qy       7801 CTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCCCTTTATGACATT 7860
             |||||||||||||||||||||||||||||||||||||||||| |||||||||||||| ||
Db       8141 CTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAAATGGCCCTTTATGATATC 8200

Qy       7861 ACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAGTATTCCCCCGCT 7920
             |||||||| |||||||||||||||||||||||||||||||||||||| |||||||||
Db       8201 ACACAAAAGCTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAGTACTCCCCCGCT 8260

Qy       7921 CAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCTATGGGTTTTTCG 7980
             |||||||| |||||||||||||| |||||||||||||| |||||||||||||||||||
Db       8261 CAGCGGGTGGAGTTTCTCTTGAAGGCATGGGCGGAAAAGAAAGACCCTATGGGTTTTTCG 8320

Qy       7981 TATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGGACTGAGGAGTCC 8040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8321 TATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGGACTGAGGAGTCC 8380

Qy       8041 ATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACACTCGCTAACTGAG 8100
             |||||||||||  || |||||||||||||||||||||||||||||||||  || ||||||
Db       8381 ATATATCGGGCTTGTTCCTTGCCCGAGGAGGCCCACACTGCCATACACTCACTGACTGAG 8440

Qy       8101 AGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGCGGGTACAGGCGT 8160
             |||||||||||||||||||| |||||||||||||||||||| ||||||||||||||||||||
Db       8441 AGACTTTACGTGGGAGGGCCCATGTTCAACAGCAAGGGCCAGACCTGCGGGTACAGGCGT 8500

Qy       8161 TGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACATGCTACGTGAAA 8220
             |||||||||||||||||||| |||||||||||||||||||||||||||||||| ||||||
Db       8501 TGCCGCGCCAGCGGGGTGCTTACCACTAGCATGGGGAACACCATCACATGCTATGTGAAA 8560

Qy       8221 GCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTGGTATGCGGCGAT 8280
             |||||||||||  ||||| |||||||||||||| ||||||||||||||||||||||||||
Db       8561 GCCTTAGCGGCCTGTAAGGCTGCAGGGATAATTGCGCCCACAATGCTGGTATGCGGCGAT 8620

Qy       8281 GACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGGAACCTGAGAGCC 8340
             |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       8621 GACTTGGTTGTCATCTCAGAGAGCCAGGGGACCGAGGAGGACGAGCGGAACCTGAGAGCC 8680

Qy       8341 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCCAGACCGGAGTAT 8400
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db       8681 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCCAGACCGGAATAT 8740

Qy       8401 GATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTGGGCCCACAAGGC 8460
             || ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db       8741 GACCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGTTGGGCCCACAAGGC 8800

Qy       8461 CGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGGGCTGCCTGGGAA 8520
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8801 CGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGGGCTGCCTGGGAA 8860

Qy       8521 ACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAGTACGCCCCGACC 8580
             |||||||||||||||||||||||||||||||| |||||||||||||||||||||| |||
Db       8861 ACAGTTAGACACTCCCCTGTCAATTCATGGCTAGGAAACATCATCCAGTACGCCCCAACC 8920

Qy       8581 ATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATGGCTCAAGACACG 8640
             |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||| ||
Db       8921 ATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATGGCCCAAGATACT 8980

Qy       8641 CTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTGAGTCCCTTGGAC 8700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8981 CTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTGAGTCCCTTGGAC 9040

Qy       8701 CTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTGCACACATACACT 8760
             |||||||| |||||||||||||||| |||||||||||||| |||||||||||||||||||
Db       9041 CTCCCAGCCATAATTGAAAGGTTACACGGGCTTGACGCTTTCTCTCTGCACACATACACT 9100

Qy       8761 CCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGA 8820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9101 CCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGA 9160

Qy       8821 GCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGTGGGGGGAGAGCG 8880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9161 GCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGTGGGGGGAGAGCG 9220

Qy       8881 GCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 8940
             |||||||||||| |||||||||||||| ||||||||||||||||||||||||||||||||
Db       9221 GCCGTTTGCGGCCGATATCTCTTCAACTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 9280

Qy       8941 TTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGCGCCGGCGGGGGC 9000
             |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db       9281 TTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACTGTCGGCGCCGGCGGGGGC 9340

Qy       9001 GACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTTGGCCTACTCCTA 9060
             |||||||||||||||||||||||||||||||||||||||| ||| ||||||||||||||
Db       9341 GACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTACTCCTTGGCCTACTCCTA 9400

Qy       9061 CTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCGG 9099
             |||||||||||||||||||||||||||||||||||||||
Db       9401 CTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCGG 9439
```

RESULT 3
US-07-925-695-7
; Sequence 7, Application US/07925695
; Patent No. 5428145
; GENERAL INFORMATION:
; APPLICANT: OKAMOTO, Hiroaki
; APPLICANT: NAKAMURA, Tetsuo
; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOME,
; TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, ANTIGEN, ANTIBODY AND
; TITLE OF INVENTION: DETECTION SYSTEMS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Beveridge, DeGrandi, Weilacher & Young
; STREET: 1850 M Street, N.W., Suite 800
; CITY: Washington
; STATE: D.C.
; COUNTRY: US
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/925,695
; FILING DATE: 19920807
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 287402/91
; FILING DATE: 09-AUG-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 360441/91
; FILING DATE: 05-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Weilacher, Robert G.
; REGISTRATION NUMBER: 20,531
; REFERENCE/DOCKET NUMBER: 06/87-48009
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 659-2811
; TELEFAX: (202) 659-1462

```
;      TELEX:  WUI 64470
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9511 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-07-925-695-7

  Query Match             62.1%;  Score 5653.8;  DB 1;  Length 9511;
  Best Local Similarity   75.5%;  Pred. No. 0;
  Matches 6868;  Conservative  104;  Mismatches 2127;  Indels    0;  Gaps    0;

Qy          1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||| ||||||||
Db        342 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACAAACCGCCGCCCACAG 401

Qy         61 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 120
              |||||||||||| ||||| ||||| |||||||||||||||||| |||||| ||||||||||
Db        402 GACGTTAAGTTCCCGGGTGGCGGTCAGATCGTTGGCGGAGTTTACTTGCTGCCGCGCAGG 461

Qy        121 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 180
              |||||||||||||||||||||||||||||||||||||||||||||| ||||| ||||||||||
Db        462 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCYGAGCGATCCCAGCCGCGTGGA 521

Qy        181 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 240
              | |||||||||||||| |||||||||||||||||| ||||| |||||||||| ||||||
Db        522 CGACGCCAGCCCATCCCGAAAGATCGGCGCTCCACCGGCAAGTCCTGGGGAAAGCCAGGA 581

Qy        241 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 300
              || || |||||||| ||||| || |||||   ||||||||| || ||||||||||||||||
Db        582 TATCCTTGGCCCCTGTACGGAAACGAGGGTTGCGGCTGGGCGGGTTGGCTCCTGTCCCCC 641

Qy        301 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 360
              || || || |||||  ||||||||||||  ||||||||||||| || |||||   ||||
Db        642 CGCGGGTCTCGTCCTACTTGGGGCCCCACCGACCCCCGGCATAGATCACGCAATTTGGGC 701

Qy        361 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 420
              |  |||||||||||| | |||||| || ||||||||||||||||||||||||||||||||
Db        702 AGAGTCATCGATACCATTACGTGTGGTTTTGCCGACCTCATGGGGTACATCCCTGTCGTT 761

Qy        421 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 480
              ||||||||| |:|| |||||||||||||||||| || || || || || ||||||||||||
Db        762 GGCGCCCCGGTYGGAGGCGTCGCCAGAGCTCTGGCACACGGTGTTAGGGTCCTGGAGGAC 821

Qy        481 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 540
              ||| | ||||  |||||||||||| |||||||||||||||| ||||||||||| |||||  ||
Db        822 GGGATAAATTACGCAACAGGGAATTTACCCGGTTGCTCTTTTTCTATCTTTTTGCTTGCT 881

Qy        541 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 600
              || ||||| ||| |||| :  || || || || |  ||||| | ||||||:||| |  ||
Db        882 CTTCTGTCATGCGTCACARTGCCAGTGTCTGCAGTGGAAGTCAGGAACATYAGTTCTAGC 941

Qy        601 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 660
              |||   |  ||||| || ||| | ||  ||||||| ||||||||||||   |  || ||
Db        942 TACTACGCCACTAATGATTGCTCAAACAACAGCATCACCTGGCAGCTCACTGACGCAGTT 1001

Qy        661 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 720
              |||||  | || ||  |||||||| || |||||  |          | |  |: |||||||||
Db       1002 CTCCATCTTCCTGGATGCGTCCCATGTGAGAAYGATAAYGGCACCTTGCRTTGCTGGATA 1061

Qy        721 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 780
              |  ||  |||| :| ||||| ||| | |  | ||| || ||||| |   || |||| ||
Db       1062 CAAGTAACACCCRACGTGGCTGTGAAACACCGCGGTGCGCTCACTCGTAGCCTGCGAACA 1121

Qy        781 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 840
              ||| |||||||| | || ||| | || |||  |||||| ||| | || |||||  ||  |
Db       1122 CACGTCGACATGATCGTAATGGCAGCTACGGCCTGCTCGGCCTTGTATGTGGGAGATGTG 1181

Qy        841 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 900
              |||||  |  |||||| |:  | | ||     ||||| || || || || | ||||   ||
Db       1182 TGCGGGGCCGTGATGATYCTATCGCAGGCTTTCATGGTATCACCACAACGCCACAACTTC 1241

Qy        901 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 960
              |||||| ||||| || |||||||||| |||  |||||| || || ||||||||||||
Db       1242 ACCCAAGAGTGCAACTGTTCCATCTACCAAGGTCACATCACCGGCCATCGCATGGCATGG 1301

Qy        961 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1020
              |||||||| |:|:|||||| || ||   |:|||||||| | || |||||    ||:||
Db       1302 GACATGATGCTRARCTGGTCTCCAACTCTTRCCATGATCCTCGCCTACGCYGCTCGYGTT 1361

Qy       1021 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1080
              |||||: |  |  | || ||:||:  ||| |  |||||||| ||  :|||:||||||:|
Db       1362 CCCGARCTGGTCCTCGAAATYATYTTCGGCGGCCATTGGGGTGTGGYGTTYGGCTTGGSC 1421

Qy       1081 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1140
              || ||||| |||||:|||||||||| ||||||:|:| |||||| ||  | |  || ||
Db       1422 TATTTCTCCATGCARGGAGCGTGGGCCAAAGTCRTYGCCATCCTCCTTCTTGTTGCGGGA 1481

Qy       1141 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1200
              ||||| ||   |||| || |    ||        || | |  :|| :|: : |  || |
Db       1482 GTGGATGCAWCCACCTATTCCASCGGYCAGSAAGCGGGTCGTRCCGYCKMKGGGWTCKCT 1541

Qy       1201 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1260
              :|| | ||| : |  ||  ||| |||||  ||:|  |  | ||||||||||||||||||
Db       1542 RGCCTCTTTAMTACTGGTGCCAAGCAGAACCTCYATTTRATCAACACCAATGGCAGCTGG 1601

Qy       1261 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1320
              ||||| |||||  |||||||||  |:| :|  || || ||  :|||| ||
Db       1602 CACATAAACCGGACTGCCCTCAATTGCAATGACAGCYTASAGACGGGTTTCMTCGCTTCC 1661

Qy       1321 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1380
              :||:| |||::||:||: ||||||   || || || ||||| ||| ||||  ||||||||
Db       1662 YTGKTTTACWMCCRCARGTTCAACAGCTCTGGCTGCCCCGAGCGCTTGTCTTCCTGCCGC 1721

Qy       1381 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1440
              |  | || | :||:||  | || |||||  |||||  |||| ||    || ||||||||
Db       1722 GGGCTGGACGAYTTYCGCATCGGCTGGGGAACCTTGGAATACGAAACCAACGTCACCAAC 1781

Qy       1441 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1500
              |:||| |||||  || || |||||||| |||||  || ||||   || ||| | ||| |
Db       1782 GATGRGGACATGAGGCCGTACTGCTGGCATTACCCCCCGAGGCCTTGCGGCATCGTCCCG 1841

Qy       1501 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1560
              || | ||| || || || || || || ||:|||||||| |||||  || || ||||||||
Db       1842 GCTAGGACGGTTTGCGGACCGGTCTATTGYTTCACCCCTAGCCCTGTTGTCGTGGGCACC 1901

Qy       1561 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1620
              || |||| ||  || |  ||||| ||||| |||||:|| || ||||| ||||||||||
Db       1902 ACTGACAAGCAGGGCGTACCCACCTACACCTGGGGRGAAAACGAGACCGATGTCTTCCTG 1961

Qy       1621 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 1680
              |:|| |||||  |||| ||||  |  | ||||||||||||||  ||||||||    ||
Db       1962 CTRAATAGCACAAGACCCCCGCGAGGAGCTTGGTTCGGCTGCACYTGGATGAACGGGACT 2021

Qy       1681 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 1740
              || | ||| ||||| ||||| |||||||| ||||| | |||     |||| ||||  ||
Db       2022 GGGTTCACTAAGACATGCGGTGCACCACCTTGCCGCATTAGGAAAGACTACAACAGCACT 2081

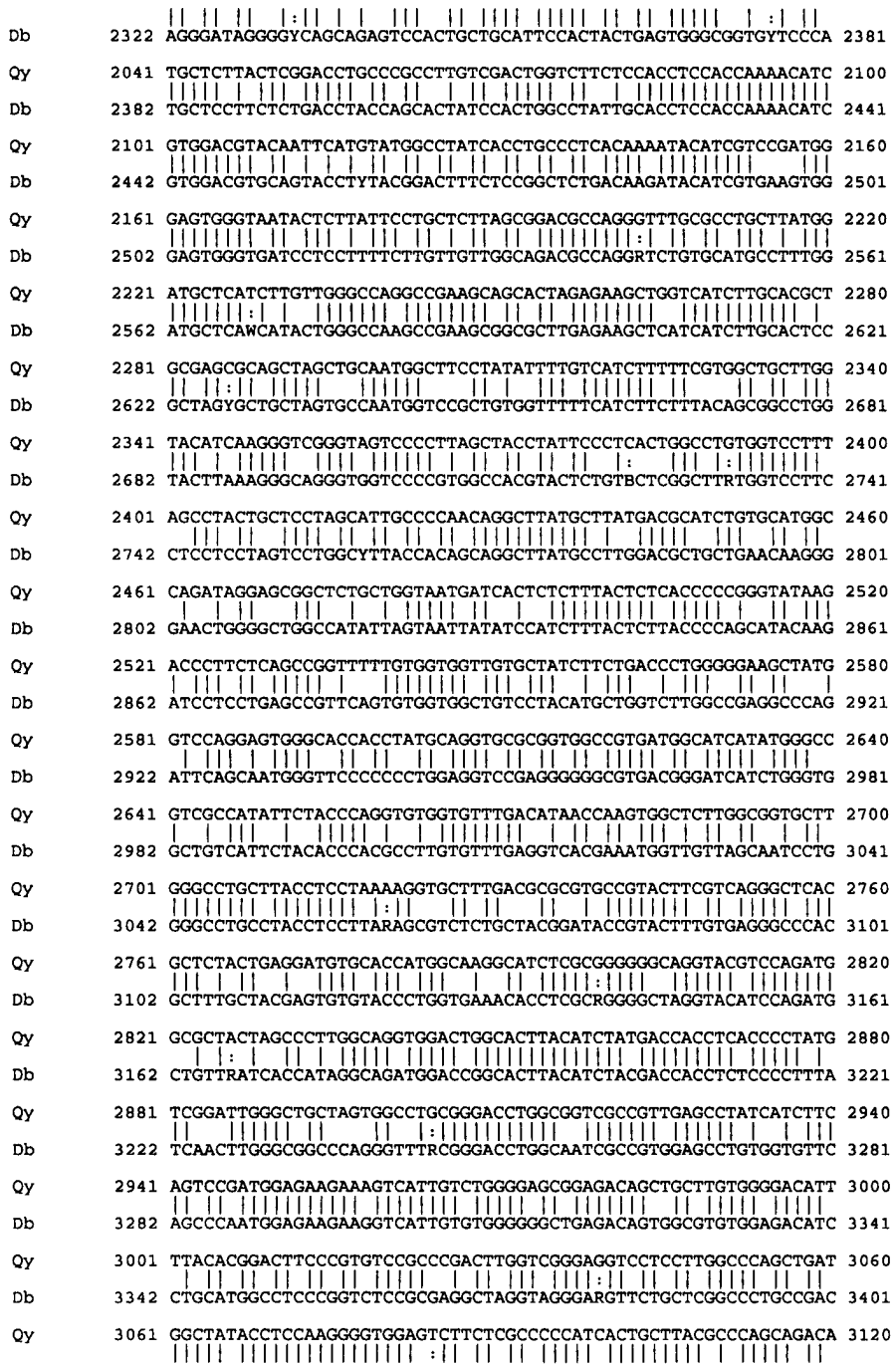
Qy       1741 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 1800
              ||  | |||||||||||| |||||||||||||||||| || ||| | || || || ||
Db       2082 CTCGATTTATTGTGCCCCACAGACTGTTTTAGGAAGCACCCAGATGCTACCTATCTTAAG 2141

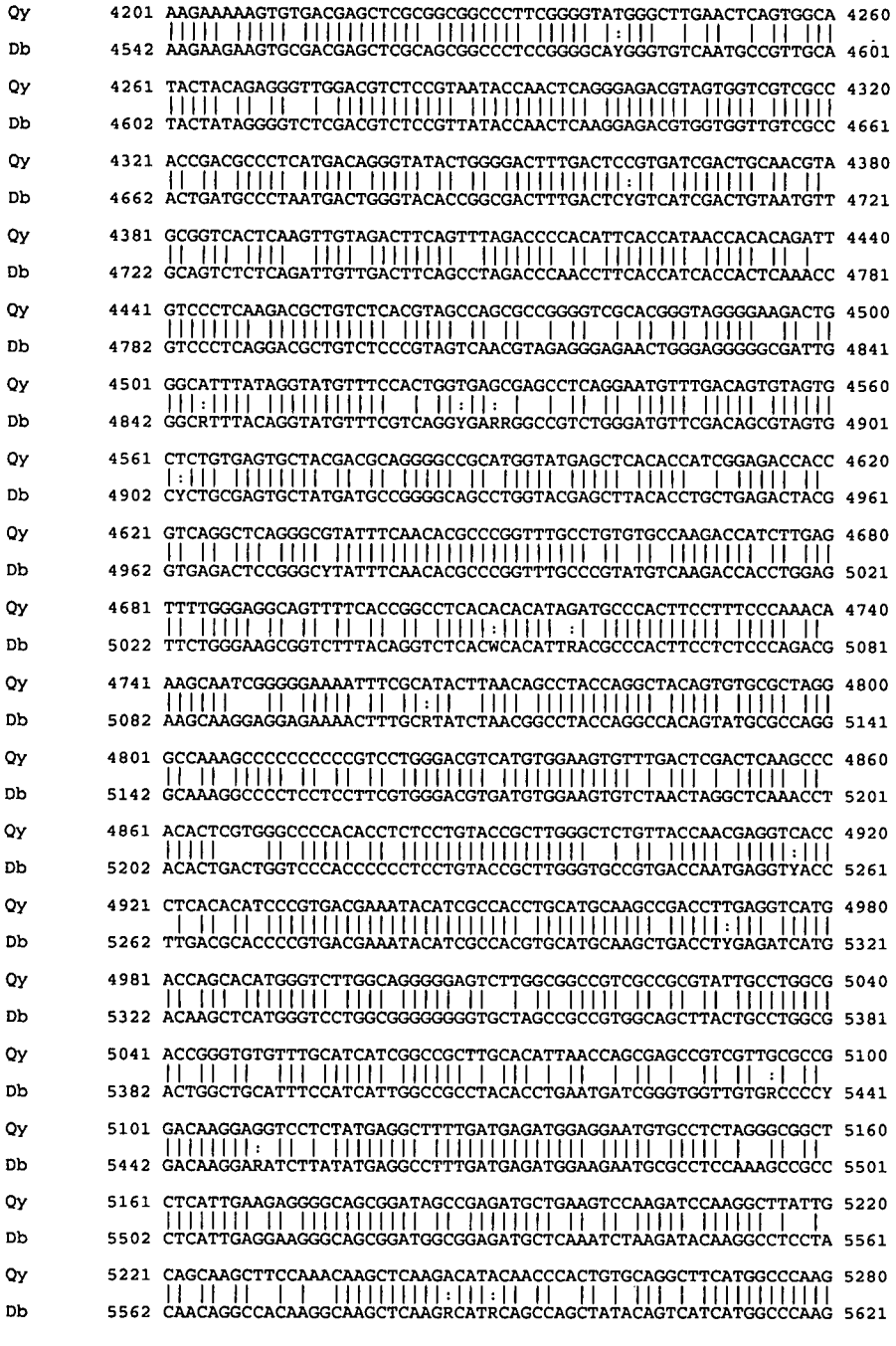
Qy       1801 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 1860
              || ||  | ||||| || || ||||||||| | |||||||| || ||::| |||
Db       2142 TGTGGAGCAGGGCCTTGGTTAACTCCCAGGTGCCTGGTAGACTACCCTTATAGRYTGTGG 2201

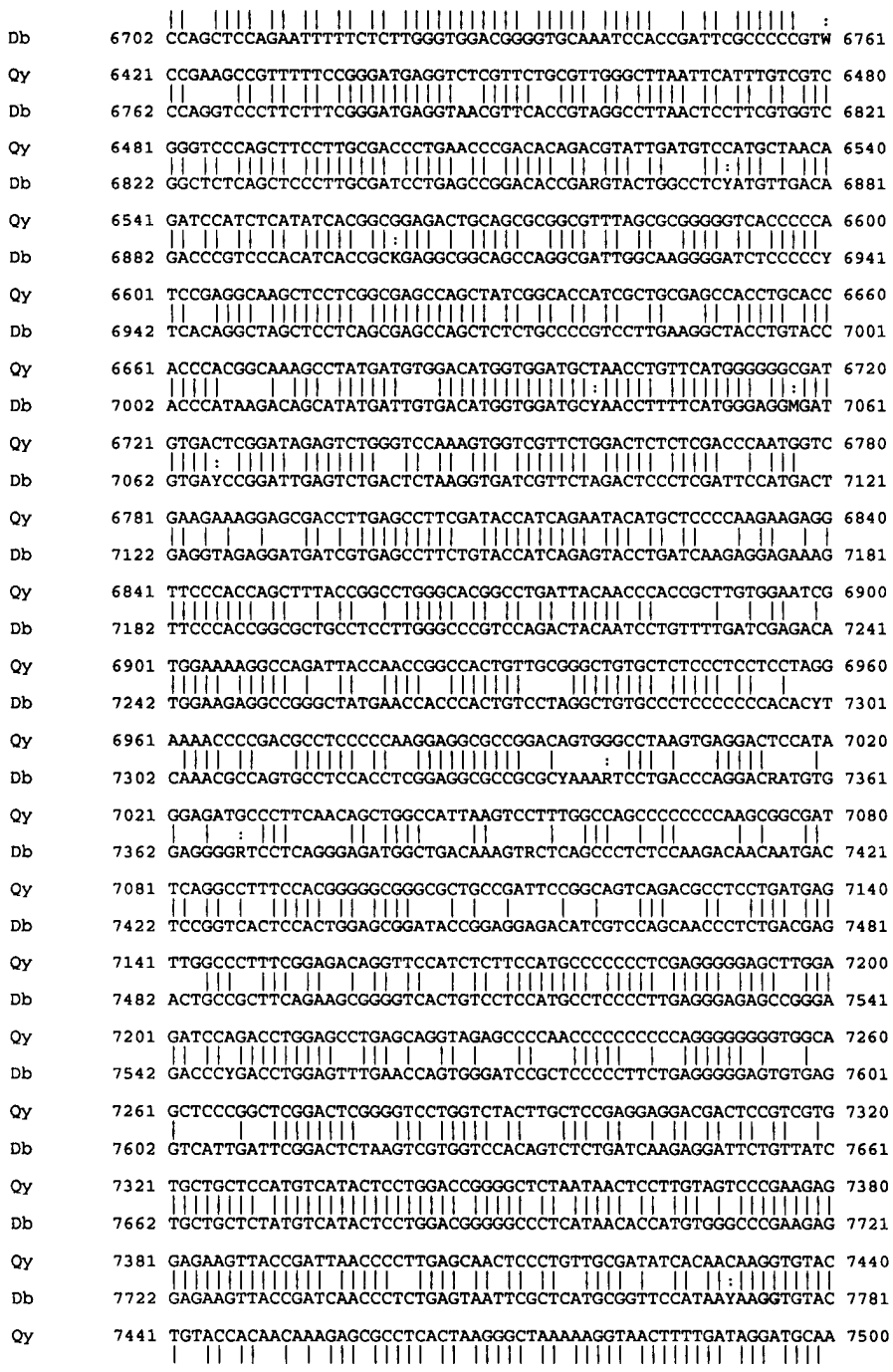
Qy       1861 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 1920
              ||||| || ||||| || ||||  ||||||||:||      |||||||||| ||||||||
Db       2202 CATTATCCGTGCACTGTAAACTTCACCATCTTYAAGGCGCGGATGTATGTAGGAGGGGTG 2261

Qy       1921 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 1980
              |||||  |  || | || |||||||| ||||| || || ||||| ||||   |||| ||
Db       2262 GAGCATCGATTCTCCGCAGCATGCAACTTCACGCGCGGAGATCGCTGCAGACTGGAAGAT 2321

Qy       1981 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2040
```


Db 2322 AGGGATAGGGGYCAGCAGAGTCCACTGCTGCATTCCACTACTGAGTGGGCGGTGYTCCCA 2381
Qy 2041 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2100
Db 2382 TGCTCCTTCTCTGACCTACCAGCACTATCCACTGGCCTATTGCACCTCCACCAAAACATC 2441
Qy 2101 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2160
Db 2442 GTGGACGTGCAGTACCTYTACGGACTTTCTCCGGCTCTGACAAGATACATCGTGAAGTGG 2501
Qy 2161 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2220
Db 2502 GAGTGGGTGATCCTCCTTTTCTTGTTGTTGGCAGACGCCAGGRTCTGTGCATGCCTTTGG 2561
Qy 2221 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2280
Db 2562 ATGCTCAWCATACTGGGCCAAGCCGAAGCGGCGCTTGAGAAGCTCATCATCTTGCACTCC 2621
Qy 2281 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATATTTTGTCATCTTTTTCGTGGCTGCTTGG 2340
Db 2622 GCTAGYGCTGCTAGTGCCAATGGTCCGCTGTGGTTTTTCATCTTCTTTACAGCGGCCTGG 2681
Qy 2341 TACATCAAGGGTCGGGTAGTCCCCTTAGCTACCTATTCCCTCACTGGCCTGTGGTCCTTT 2400
Db 2682 TACTTAAAGGGCAGGGTGGTCCCCGTGGCCACGTACTCTGTBCTCGGCTTRTGGTCCTTC 2741
Qy 2401 AGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTGCATGGC 2460
Db 2742 CTCCTCCTAGTCCTGGCYTTACCACAGCAGGCTTATGCCTTGGACGCTGCTGAACAAGGG 2801
Qy 2461 CAGATAGGAGCGGCTCTGCTGGTAATGATCACTCTCTTTACTCTCACCCCCGGGTATAAG 2520
Db 2802 GAACTGGGGCTGGCCATATTAGTAATTATATCCATCTTTACTCTTACCCCAGCATACAAG 2861
Qy 2521 ACCCTTCTCAGCCGGTTTTTGTGGTGGTTGTGCTATCTTCTGACCCTGGGGGAAGCTATG 2580
Db 2862 ATCCTCCTGAGCCGTTCAGTGTGGTGGCTGTCCTACATGCTGGTCTTGGCCGAGGCCCAG 2921
Qy 2581 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGCATCATATGGGCC 2640
Db 2922 ATTCAGCAATGGGTTCCCCCCCTGGAGGTCCGAGGGGGGCGTGACGGGATCATCTGGGTG 2981
Qy 2641 GTCGCCATATTCTACCCAGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 2700
Db 2982 GCTGTCATTCTACACCCACGCCTTGTGTTTGAGGTCACGAAATGGTTGTTAGCAATCCTG 3041
Qy 2701 GGGCCTGCTTACCTCCTAAAAGGTGCTTTGACGCGCGTGCCGTACTTCGTCAGGGCTCAC 2760
Db 3042 GGGCCTGCCTACCTCCTTARAGCGTCTCTGCTACGGATACCGTACTTTGTGAGGGCCCAC 3101
Qy 2761 GCTCTACTGAGGATGTGCACCATGGCAAGGCATCTCGCGGGGGGCAGGTACGTCCAGATG 2820
Db 3102 GCTTTGCTACGAGTGTGTACCCTGGTGAAACACCTCGCRGGGGCTAGGTACATCCAGATG 3161
Qy 2821 GCGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 2880
Db 3162 CTGTTRATCACCATAGGCAGATGGACCGGCACTTACATCTACGACCACCTCTCCCCTTTA 3221
Qy 2881 TCGGATTGGGCTGCTAGTGGCCTGCGGGACCTGGCGGTCGCCGTTGAGCCTATCATCTTC 2940
Db 3222 TCAACTTGGGCGGCCCAGGGTTTRCGGGACCTGGCAATCGCCGTGGAGCCTGTGGTGTTC 3281
Qy 2941 AGTCCGATGGAGAAGAAAGTCATTGTCTGGGGAGCGGAGACAGCTGCTTGTGGGGACATT 3000
Db 3282 AGCCCAATGGAGAAGAAGGTCATTGTGTGGGGGGCTGAGACAGTGGCGTGTGGAGACATC 3341
Qy 3001 TTACACGGACTTCCCGTGTCCGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3060
Db 3342 CTGCATGGCCTCCCGGTCTCCGCGAGGCTAGGTAGGGARGTTCTGCTCGGCCCTGCCGAC 3401
Qy 3061 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3120
Db 3402 GGCTACACCTCCAAGGGGTGGAAKCTCCTAGCTCCCATTACTGCTTACACTCAGCAAACT 3461
Qy 3121 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCT 3180
Db 3462 CGTGGTCTCCTGGGTGCTATCGTGGTCAGCCTAACGGGCCGCGACAAAAATGAGCAGGCT 3521
Qy 3181 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3240
Db 3522 GGGCAGGTCCAGGTTCTGTCCTCCGTCACACAAACTTTCTTGGGGACATCCATTTCGGGC 3581
Qy 3241 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3300
Db 3582 GTCCTCTGGACAGTATATCACGGGGCTGGTAATAAGACCTTGGCCGGCCCCAAGGGACCA 3641
Qy 3301 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGG 3360
Db 3642 GTCACTCAGATGTACACCAGCGCAGAAGGGGACCTCGTGGGATGGCCTAGTCCCCCCGGG 3701
Qy 3361 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3420
Db 3702 ACTAAGTCATTGGACCCCTGTACCTGCGGGGCCGTAGACCTCTACCTGGTCACCCGAAAC 3761
Qy 3421 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3480
Db 3762 GCTGATGTCATTCCGGTCCGGAGGAAAGATGACCGACGGGGTGCATTACTCTCGCCAAGG 3821
Qy 3481 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3540
Db 3822 CCCCTCTCAACCCTCAAAGGATCATCCGGAGGGCCCGTGCTCTGCTCWAGGGGACACGCC 3881
Qy 3541 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3600
Db 3882 GTGGGCTTGTTCAGAGCGGCCGTGTGTGCCAGGGGTGTAGCCAAATCTATTGACTTCATC 3941
Qy 3601 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 3660
Db 3942 CCCGTCGAATCACTCGATRTCGCCACACGGACGCCCAGTTTCTCTGACAACAGTRCGCCG 4001
Qy 3661 CCTGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 3720
Db 4002 CCAGCTGTGCCCCAGTCTTACCAGGTGGGTTACTTGCACGCACCAACAGGCAGCGGAAAG 4061
Qy 3721 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 3780
Db 4062 AGCACCAAGGTCCCTGCCGCGTATGCCAGTCAGGGGTATAAAGTACTCGTACTAAATCCC 4121
Qy 3781 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 3840
Db 4122 TCTGTCGCGGCCACACTTGGTTTTGGGGCCTACATGTCCAAAGCCCACGGGATCAACCCT 4181
Qy 3841 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 3900
Db 4182 AATATCAGAACTGGAGTGCGGACCGTTACCACCGGGGACTCTATCACTTACTCCACTTAT 4241
Qy 3901 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 3960
Db 4242 GGCAAGTTTATCGCAGATGGAGGCTGTGCAGCCGGTGCCTATGACATCATCATATGCGAC 4301
Qy 3961 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4020
Db 4302 GAATGCCATTCAGTGGACGCTACTACCATCCTTGGCATTGGAACAGTCCTTGACCAAGCT 4361
Qy 4021 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4080
Db 4362 GAGACCGCAGGCGTCAGGCTAGTGGTYTTGGCCACAGCCACGCCTCCCGGTACGGTGACA 4421
Qy 4081 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4140
Db 4422 ACTCCCCACAGTAACATAGAGGAGGTGGCCCTTGGTCACGAGGGCGAGATCCCTTTTTAT 4481
Qy 4141 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4200
Db 4482 GGCAAAGCTATTCCCCTAGCTTTCATCAAGGGGGGCAGACACTTGATCTTTTGCCATTCA 4541


Qy 4201 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4260
Db 4542 AAGAAGAAGTGCGACGAGCTCGCAGCGGCCCTCCGGGGCAYGGGTGTCAATGCCGTTGCA 4601
Qy 4261 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4320
Db 4602 TACTATAGGGGTCTCGACGTCTCCGTTATACCAACTCAAGGAGACGTGGTGGTTGTCGCC 4661
Qy 4321 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4380
Db 4662 ACTGATGCCCTAATGACTGGGTACACCGGCGACTTTGACTCYGTCATCGACTGTAATGTT 4721
Qy 4381 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4440
Db 4722 GCAGTCTCTCAGATTGTTGACTTCAGCCTAGACCCAACCTTCACCATCACCACTCAAACC 4781
Qy 4441 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4500
Db 4782 GTCCCTCAGGACGCTGTCTCCCGTAGTCAACGTAGAGGGAGAACTGGGAGGGGGCGATTG 4841
Qy 4501 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4560
Db 4842 GGCRTTTACAGGTATGTTTCGTCAGGYGARRGGCCGTCTGGGATGTTCGACAGCGTAGTG 4901
Qy 4561 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4620
Db 4902 CYCTGCGAGTGCTATGATGCCGGGGCAGCCTGGTACGAGCTTACACCTGCTGAGACTACG 4961
Qy 4621 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 4680
Db 4962 GTGAGACTCCGGGCYTATTTCAACACGCCCGGTTTGCCCGTATGTCAAGACCACCTGGAG 5021
Qy 4681 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 4740
Db 5022 TTCTGGGAAGCGGTCTTTACAGGTCTCACWCACATTRACGCCCACTTCCTCTCCCAGACG 5081
Qy 4741 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 4800
Db 5082 AAGCAAGGAGGAGAAAACTTTGCRTATCTAACGGCCTACCAGGCCACAGTATGCGCCAGG 5141
Qy 4801 GCCAAAGCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 4860
Db 5142 GCAAAGGCCCCTCCTCCTTCGTGGGACGTGATGTGGAAGTGTCTAACTAGGCTCAAACCT 5201
Qy 4861 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 4920
Db 5202 ACACTGACTGGTCCCACCCCCCTCCTGTACCGCTTGGGTGCCGTGACCAATGAGGTYACC 5261
Qy 4921 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 4980
Db 5262 TTGACGCACCCCGTGACGAAATACATCGCCACGTGCATGCAAGCTGACCTYGAGATCATG 5321
Qy 4981 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5040
Db 5322 ACAAGCTCATGGGTCCTGGCGGGGGGGGTGCTAGCCGCCGTGGCAGCTTACTGCCTGGCG 5381
Qy 5041 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5100
Db 5382 ACTGGCTGCATTTCCATCATTGGCCGCCTACACCTGAATGATCGGGTGGTTGTGRCCCCY 5441
Qy 5101 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5160
Db 5442 GACAAGGARATCTTATATGAGGCCTTTGATGAGATGGAAGAATGCGCCTCCAAAGCCGCC 5501
Qy 5161 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5220
Db 5502 CTCATTGAGGAAGGGCAGCGGATGGCGGAGATGCTCAAATCTAAGATACAAGGCCTCCTA 5561
Qy 5221 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5280
Db 5562 CAACAGGCCACAAGGCAAGCTCAAGRCATRCAGCCAGCTATACAGTCATCATGGCCCAAG 5621
Qy 5281 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5340
Db 5622 CTTGAACAATTTTGGGCCAAACACATGTGGAACTTCATCAGTGGTATACAGTACCTAGCA 5681
Qy 5341 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5400
Db 5682 GGACTCTCCACCCTACCGGGAAATCCTGCAGTRGCATCAATGATGGCTTTTAGCGCCGCG 5741
Qy 5401 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5460
Db 5742 CTGACTAGCCCACTACCCACCAGCACCACCATCCTCTTGAACATCATGGGAGGATGCTTG 5801
Qy 5461 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5520
Db 5802 GCCTCYCAGATTGCCCCCCCTGCCGGAGCCACYGGCTTCGTTGTCAGTGGTCTAGTGGGG 5861
Qy 5521 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5580
Db 5862 GCGGCCGTCGGAAGCATAGGCCTGGGTAAGATACTGGTGGACGTTTTGGCCGGGTACGGC 5921
Qy 5581 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5640
Db 5922 GCAGGCATTTCAGGGGCCCTCGTAGCTTTTAAGATCATGAGCGGCGAGAAGCCCACGGTA 5981
Qy 5641 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 5700
Db 5982 GAAGACGTTGTGAATCTCCTGCCTGCTATYCTGTCTCCTGGTGCGYTGGTAGTGGGAGTC 6041
Qy 5701 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 5760
Db 6042 ATCTGTGCAGCAATYCTGCGCCGCCACGTCGGTCAGGGAGAGGGRGCGGTCCAGTGGATG 6101
Qy 5761 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 5820
Db 6102 AACAGACTGATCGCCTTCGCCTCCAGGGGAAACCACGTTGCCCCTACCCACTACGTGGTG 6161
Qy 5821 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 5880
Db 6162 GAGTCTGACGCTTCACAGCGTGTRACGCAGGTGCTGAGTTCACTTACAATTACCAGCTTA 6221
Qy 5881 CTCAGAAGACTCCACAACTGGATTACTGAGGACTGCCCCATCCCATGCGGCGGCTCGTGG 5940
Db 6222 CTTAGGAGACTACATGCCTGGATCACTGAAGATTGCCCARTCCCATGCTCGGGGTCTTGG 6281
Qy 5941 CTCCGCGATGTGTGGGACTGGGTTTGCACCATCCTAACAGACTTTAAAAATTGGCTGACC 6000
Db 6282 CTCCAGGACATTTGGGATTGGGTTTGTTCCATCCTCACAGACTTYAAAAACTGGCTGTCT 6341
Qy 6001 TCCAAATTATTCCCAAAGATGCCCGGCCTCCCCTTTGTCTCCTGTCAAAAGGGGTACAAG 6060
Db 6342 TCAAAATTACTCCCCAAGATGCCCGGCATTCCCTTTATCTCTTGCCAGAAGGGATACAAG 6401
Qy 6061 GGCGTGTGGGCCGGCACTGGCATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6120
Db 6402 GGTGTATGGGCTGGTACGGGTGTCATGACYACTCGRTRCCCATGTGGAGCAAACATCTCG 6461
Qy 6121 GGCAATGTCCGCTTGGGCTCCATGAGAATCACGGGGCCTAAGACCTGCATGAATATCTGG 6180
Db 6462 GGCCATGTCCGCATGGGCACCATGAAAATAACAGGCCCGAAGACTTGCTTGAACCTGTGG 6521
Qy 6181 CAGGGGACCTTTCCTATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCGCCA 6240
Db 6522 CAGGGGACTTTCCCCATTAATTGTTACACAGAAGGGCCYTGCGTGCCAAAACCCCCTCCT 6581
Qy 6241 AACTTTAAGGTCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6300
Db 6582 AATTACAAGACCGCAATTTGGAGGGTGGCAGCGTCGGAGTACGTTGAGGTCACACAGCAT 6641
Qy 6301 GGGTCATACCACTACATAACAGGACTCACCACTGATAACTTGAAAGTCCCCTGCCAACTA 6360
Db 6642 GGCTCTTTCTCGTATGTAACRGGGTTAACCAGTGACAACCTTAAGGTYCCTTGCCAGGTA 6701
Qy 6361 CCCTCTCCCGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCACA 6420


Db 6702 CCAGCTCCAGAATTTTTCTCTTGGGTGGACGGGGTGCAAATCCACCGATTCGCCCCCGTW 6761
Qy 6421 CCGAAGCCGTTTTTCCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6480
Db 6762 CCAGGTCCCTTCTTTCGGGATGAGGTAACGTTCACCGTAGGCCTTAACTCCTTCGTGGTC 6821
Qy 6481 GGGTCCCAGCTTCCTTGCGACCCTGAACCCGACACAGACGTATTGATGTCCATGCTAACA 6540
Db 6822 GGCTCTCAGCTCCCTTGCGATCCTGAGCCGGACACCGARGTACTGGCCTCYATGTTGACA 6881
Qy 6541 GATCCATCTCATATCACGGCGGAGACTGCAGCGCGGCGTTTAGCGCGGGGGTCACCCCCA 6600
Db 6882 GACCCGTCCCACATCACCGCKGAGGCGGCAGCCAGGCGATTGGCAAGGGGATCTCCCCCY 6941
Qy 6601 TCCGAGGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 6660
Db 6942 TCACAGGCTAGCTCCTCAGCGAGCCAGCTCTCTGCCCCGTCCTTGAAGGCTACCTGTACC 7001
Qy 6661 ACCCACGGCAAAGCCTATGATGTGGACATGGTGGATGCTAACCTGTTCATGGGGGGCGAT 6720
Db 7002 ACCCATAAGACAGCATATGATTGTGACATGGTGGATGCYAACCTTTTCATGGGAGGMGAT 7061
Qy 6721 GTGACTCGGATAGAGTCTGGGTCCAAAGTGGTCGTTCTGGACTCTCTCGACCCAATGGTC 6780
Db 7062 GTGAYCCGGATTGAGTCTGACTCTAAGGTGATCGTTCTAGACTCCCTCGATTCCATGACT 7121
Qy 6781 GAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCAGAATACATGCTCCCCAAGAAGAGG 6840
Db 7122 GAGGTAGAGGATGATCGTGAGCCTTCTGTACCATCAGAGTACCTGATCAAGAGGAGAAAG 7181
Qy 6841 TTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCGCTTGTGGAATCG 6900
Db 7182 TTCCCACCGGCGCTGCCTCCTTGGGCCCGTCCAGACTACAATCCTGTTTTGATCGAGACA 7241
Qy 6901 TGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTCCCTCCTCCTAGG 6960
Db 7242 TGGAAGAGGCCGGGCTATGAACCACCCACTGTCCTAGGCTGTGCCCTCCCCCCCACACYT 7301
Qy 6961 AAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGTGAGGACTCCATA 7020
Db 7302 CAAACGCCAGTGCCTCCACCTCGGAGGCGCCGCGCYAAARTCCTGACCCAGGACRATGTG 7361
Qy 7021 GGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCCCCAAGCGGCGAT 7080
Db 7362 GAGGGGRTCCTCAGGGAGATGGCTGACAAAGTRCTCAGCCCTCTCCAAGACAACAATGAC 7421
Qy 7081 TCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACGCCTCCTGATGAG 7140
Db 7422 TCCGGTCACTCCACTGGAGCGGATACCGGAGGAGACATCGTCCAGCAACCCTCTGACGAG 7481
Qy 7141 TTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCTCGAGGGGGAGCTTGGA 7200
Db 7482 ACTGCCGCTTCAGAAGCGGGGTCACTGTCCTCCATGCCTCCCCTTGAGGGAGAGCCGGGA 7541
Qy 7201 GATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCCAGGGGGGGGTGGCA 7260
Db 7542 GACCCYGACCTGGAGTTTGAACCAGTGGGATCCGCTCCCCCTTCTGAGGGGGAGTGTGAG 7601
Qy 7261 GCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGACGACTCCGTCGTG 7320
Db 7602 GTCATTGATTCGGACTCTAAGTCGTGGTCCACAGTCTCTGATCAAGAGGATTCTGTTATC 7661
Qy 7321 TGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGTAGTCCCGAAGAG 7380
Db 7662 TGCTGCTCTATGTCATACTCCTGGACGGGGGCCCTCATAACACCATGTGGGCCCGAAGAG 7721
Qy 7381 GAGAAGTTACCGATTAACCCCTTGAGCAACTCCCTGTTGCGATATCACAACAAGGTGTAC 7440
Db 7722 GAGAAGTTACCGATCAACCCTCTGAGTAATTCGCTCATGCGGTTCCATAAYAAGGTGTAC 7781
Qy 7441 TGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTTGATAGGATGCAA 7500
Db 7782 TCCACAACCTCGAGGAGTGCCTCTCTGAGGGCAAAGAAGGTGACTTTTGACAGGGTGCAG 7841
Qy 7501 GTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCGGCCTCCAAGGTC 7560
Db 7842 GTGCTGGACGCACACTATGACTCAGTCTTGCAGGACGTTAAGCGGGCCGCCTCTAAGGTT 7901
Qy 7561 ACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCCCATTCTGCAAGA 7620
Db 7902 RGTGCGAGGCTCCTCACAGTAGAGGAAGCCTGCGCGCTGACCCCGCCCCACTCCGCCAAA 7961
Qy 7621 TCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGGGCCGTTAACCAC 7680
Db 7962 TCGCGATACGGATTTGGGGCAAAAGAGGTGCGCAGCTTATCCAGGAGGGCCGTTAACCAC 8021
Qy 7681 ATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATTCCCACAACCATT 7740
Db 8022 ATCCGGTCCGTGTGGGAGGACCTCCTGGAAGACCAACRTACCCCAATTGACACAACTATC 8081
Qy 7741 ATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAGAAAGCAGCTCGC 7800
Db 8082 ATGGCTAAAAATGAGGTGTTCTGCATTGATCCAACTAARGGTGGGAAAAAGCCAGCTCGC 8141
Qy 7801 CTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCCCTTTATGACATT 7860
Db 8142 CTCATCGTATACCCCGACCTTGGGGTCAGGGTGTGCGAAAAGATGGCCCTCTATGACATC 8201
Qy 7861 ACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAGTATTCCCCCGCT 7920
Db 8202 RCACAAAAGCTTCCCAAAGCGATAATGGGGCCATCCTATGGGTTCCAATACTCTCCCGCA 8261
Qy 7921 CAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCTATGGGTTTTTCG 7980
Db 8262 GAACGGGTCGATTTCCTCCTCAAAGCTTGGGGAAGTAAGAAGGACCCAATGGGGTTCTCG 8321
Qy 7981 TATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGGACTGAGGAGTCC 8040
Db 8322 TATGACACCCGCTGCTTTGACTCAACCGTCACGGAGAGGGACATAAGAACAGAAGAATCC 8381
Qy 8041 ATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACACTCGCTAACTGAG 8100
Db 8382 ATATATCAGGCTTGTTCTCTGCCTCAAGAAGCCAGAACTGTCATACACTCGCTCACTGAG 8441
Qy 8101 AGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGCGGGTACAGGCGT 8160
Db 8442 AGACTTTACGTAGGAGGGCCCATGACAAACAGCAAAGGGCAATCCTGCGGCTACAGGCGT 8501
Qy 8161 TGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACATGCTACGTGAAA 8220
Db 8502 TGCCGCGCAAGCGGKGTTTTCACCACCAGCATGGGGAATACCATGACATGTTACATCAAA 8561
Qy 8221 GCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTGGTATGCGGCGAT 8280
Db 8562 GCCCTTGCAGCGTGTAAGGCTGCRGGGATCGTGGACCCTGTTATGTTGGTGTGTGGAGAC 8621
Qy 8281 GACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGGAACCTGAGAGCC 8340
Db 8622 GACCTGGTCGTCATCTCAGAGAGCCAAGGTAACGAGGAGGACGAGCGAAACCTGAGAGCT 8681
Qy 8341 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCCAGACCGGAGTAT 8400
Db 8682 TTCACGGAGGCTATGACCAGGTATTCCGCCCCTCCCGGTGACCTTCCCAGACCGGAATAT 8741
Qy 8401 GATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTGGGCCCACAAGGC 8460
Db 8742 GACTTGGAGCTTATAACATCCTGCTCCTCAAACGTATCGGTAGCGCTGGACTCTCGGGGT 8801
Qy 8461 CGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGGGCTGCCTGGGAA 8520
Db 8802 CGCCGCCGGTACTTCCTAACCAGAGACCCTACCACTCCAATCACCCGAGCTGCTTGGGAA 8861
Qy 8521 ACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAGTACGCCCCGACC 8580
Db 8862 ACAGTAAGACACTCCCCTGTCAATTCTTGGCTGGGCAACATCATCCAGTACGCCCCCACA 8921

```
Qy       8581 ATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATGGCTCAAGACACG 8640
              || ||||  || ||||||  | ||||| ||||||||||||||  ||  |||| || |||||
Db       8922 ATCTGGGTCCGGATGGTCATAATGACTCACTTCTTCTCCATACTATTGGCCCAGGACACT 8981

Qy       8641 CTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTGAGTCCCTTGGAC 8700
              ||| |||| || ||||| |||||||||||||| || || ||||| || | ||| || |||
Db       8982 CTGAACCAAAATCTCAATTTTGAGATGTACGGGGCAGTATACTCGGTCAATCCATTAGAC 9041

Qy       8701 CTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTGCACACATACACT 8760
              || || || |||||||||||| |||||||||||| || ||||| ||||||||||| ||
Db       9042 CTACCGGCCATAATTGAAAGGCTACATGGGCTTGAAGCCTTTTCACTGCACACATACTCT 9101

Qy       8761 CCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGA 8820
              ||||||||||  ||||||||||  || | |||||||||||||| ||||| ||||| |||
Db       9102 CCCCACGAACTCTCACGGGTGGCAGCAACTCTCAGAAAACTTGGAGCGCCTCCCCTTAGA 9161

Qy       8821 GCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGTGGGGGGAGAGCG 8880
              |||||||||||| ||||| ||||| || || || || |||||| |||  || | ||| |||
Db       9162 GCGTGGAAGAGTCGGGCGCGTGCCGTGAGAGCTTCACTCATCGCCCAAGGAGCGAGGGCG 9221

Qy       8881 GCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 8940
              ||| |||| || || || |||||||| |||||||||| || ||||||||||||||||||
Db       9222 GCCATTTGTGGCCGCTACCTCTTCAACTGGGCGGTGAAAACAAAGCTCAAACTCACTCCA 9281

Qy       8941 TTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGCGCCGGCGGGGGC 9000
              ||||| || ||| ||| |||||||||||| | ||||||||||| ||||||||||||||||
Db       9282 TTGCCCGAGGCGAGCCGCCTGGATTTATCCGGGTGGTTCACCGTGGGCGCCGGCGGGGGC 9341

Qy       9001 GACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTTGGCCTACTCCTA 9060
              |||||||||||||||||||||| |||:|||||||| |||| ||| || ||||||||||
Db       9342 GACATTTATCACAGCGTGTCGCATGCYCGACCCCGCCTATTACTCCTTTGCCTACTCCTA 9401

Qy       9061 CTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCGG 9099
              |||   ||||| |||||| | ||  ||||||||||||||
Db       9402 CTTAGCGTAGGAGTAGGCATCTTTTTACTCCCCGCTCGG 9440
```

```
RESULT 4
US-07-925-695-6
; Sequence 6, Application US/07925695
; Patent No. 5428145
;  GENERAL INFORMATION:
;    APPLICANT:  OKAMOTO, Hiroaki
;    APPLICANT:  NAKAMURA, Tetsuo
;    TITLE OF INVENTION:  NON-A, NON-B HEPATITIS VIRUS GENOME,
;    TITLE OF INVENTION:  POLYNUCLEOTIDES, POLYPEPTIDES, ANTIGEN, ANTIBODY AND
;    TITLE OF INVENTION:  DETECTION SYSTEMS
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Beveridge, DeGrandi, Weilacher & Young
;      STREET:  1850 M Street, N.W., Suite 800
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  US
;      ZIP:  20036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/925,695
;      FILING DATE:  19920807
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 287402/91
;      FILING DATE:  09-AUG-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 360441/91
;      FILING DATE:  05-DEC-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Weilacher, Robert G.
;      REGISTRATION NUMBER:  20,531
;      REFERENCE/DOCKET NUMBER:  06/87-48009
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 659-2811
;      TELEFAX:  (202) 659-1462
;      TELEX:  WUI 64470
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9511 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-07-925-695-6

  Query Match             62.1%;  Score 5651;  DB 1;  Length 9511;
  Best Local Similarity   59.4%;  Pred. No. 0;
  Matches 5403;  Conservative 1567;  Mismatches 2129;  Indels    0;  Gaps    0;

Qy          1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 60
              |:|||||||||:||:|||||:|||||||||||||||||||||||| ||||| ||||||||
Db        342 AUGAGCACAAAUCCUAAACCUCAAAGAAAAACCAAAAGAAACACAAACCGCCGCCCACAG 401

Qy         61 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 120
              ||||::|||:: ||||| ||||| ||||:||::|||||||: :||::| :|||||||||
Db        402 GACGUUAAGUUCCCGGGUGGCGGUCAGAUCGUUGGCGGAGUUUACUUGCUGCCGCGCAGG 461

Qy        121 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 180
              |||||||||::|||:|:|||||||||||||||||::| ||||| :||||||| ||:|||
Db        462 GGCCCCAGGUUGGGUGUGCGCGCGACAAGGAAGACUUCYGAGCGAUCCCAGCCGCGUGGA 521

Qy        181 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 240
              | |||||||||:||| |||||:|||||:|||| |||| :||:|||||| ||||||
Db        522 CGACGCCAGCCCAUCCCGAAAGAUCGGCGCUCCACCGGCAAGUCCUGGGGAAAGCCAGGA 581

Qy        241 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 300
              :| || :|||||: :|||| || |||||   ||||:|||| || :|||:||:|:|||||
Db        582 UAUCCUUGGCCCCUGUACGGAAACGAGGGUUGCGGCUGGGCGGGUUGGCUCCUGUCCCCC 641

Qy        301 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 360
              || || :| ||:||  |::|||||||||  ||||||||||:|| :| |||||  :|||
Db        642 CGCGGGUCUCGUCCUACUUGGGGCCCCACCGACCCCCGGCAUAGAUCACGCAAUUUGGGC 701

Qy        361 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 420
              |  |:||:|||:||| : |||:| || :::||||||||:||:||||:|||:|:||:
Db        702 AGAGUCAUCGAUACCAUUACGUGUGGUUUUGCCGACCUCAUGGGGUACAUCCCUGUCGUU 761

Qy        421 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 480
              ||||||||| ::|| ||||:|||||||||:|: || || || |: || |:||:|||||||
Db        762 GGCGCCCCGGUYGGAGGCGUCGCCAGAGCUCUGGCACACGGUGUUAGGGUCCUGGAGGAC 821

Qy        481 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 540
              ||| : ||::  |||||||||||| ::|||||::||:| ::::[:|:|:: ::||: ||
Db        822 GGGAUAAAUUACGCAACAGGGAAUUUACCCGGUUGCUCUUUUUCUAUCUUUUUGCUUGCU 881

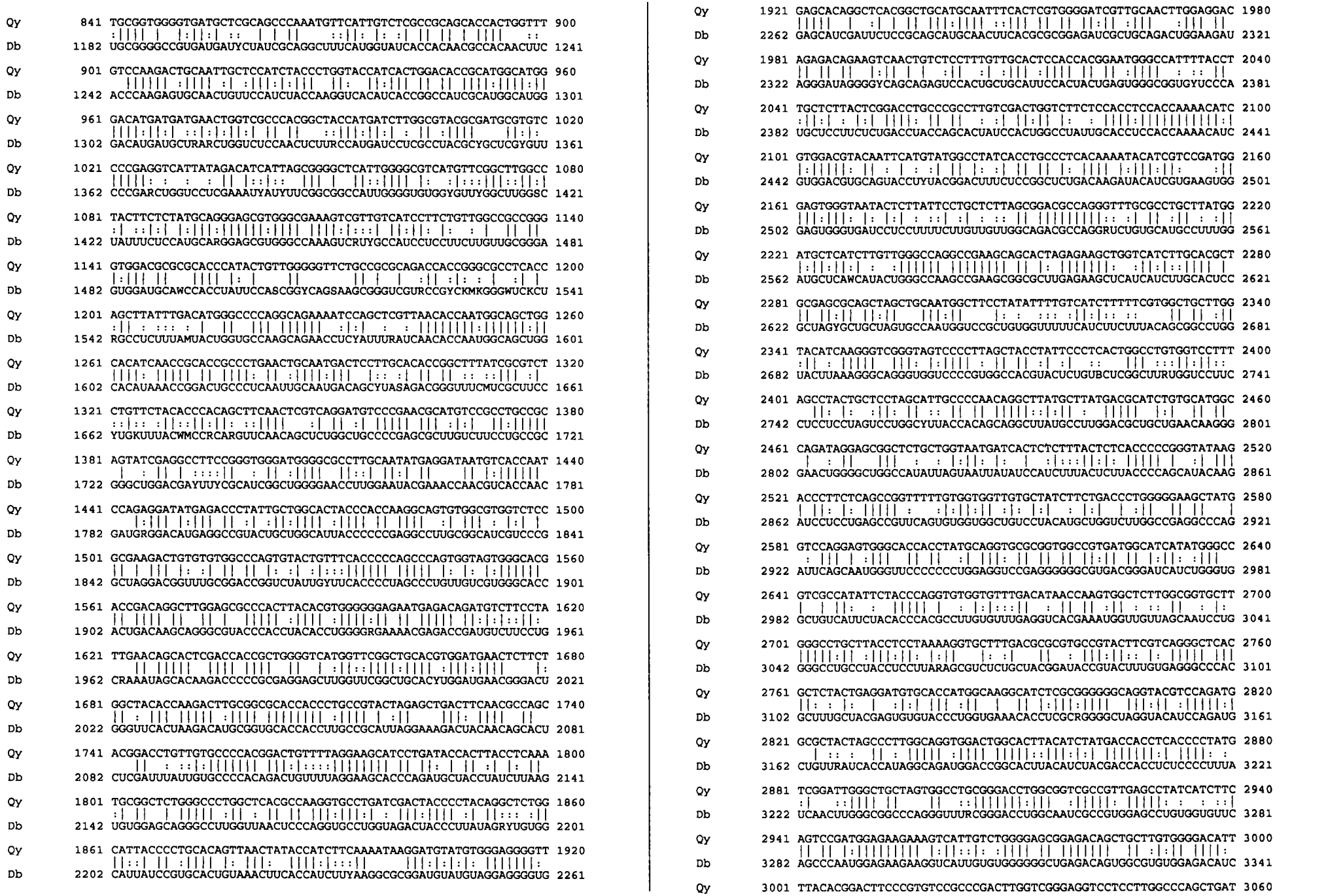
Qy        541 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 600
              |: |:|:| :|| :||| :  || |: :| || |  |||| | |||||::||: |  ||
Db        882 CUUCUGUCAUGCGUCACARUGCCAGUGUCUGCAGUGGAAGUCAGGAACAUYAGUUCUAGC 941

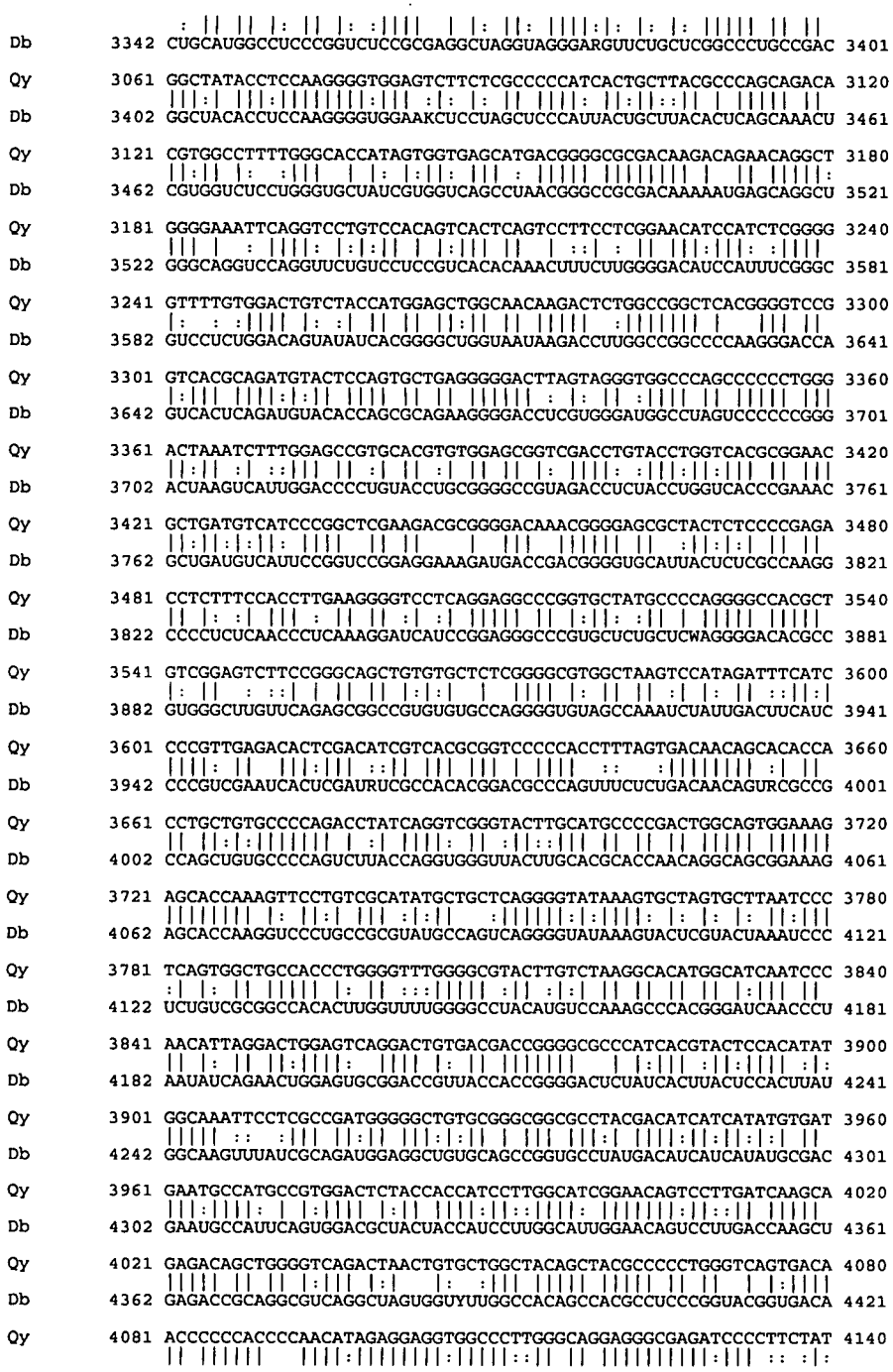
Qy        601 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 660
              :||   |  ||:|| || :|| | ||  ||||||: |||:||||||:|   |  || |:
Db        942 UACUACGCCACUAAUGAUUGCUCAAACAACAGCAUCACCUGGCAGCUCACUGACGCAGUU 1001

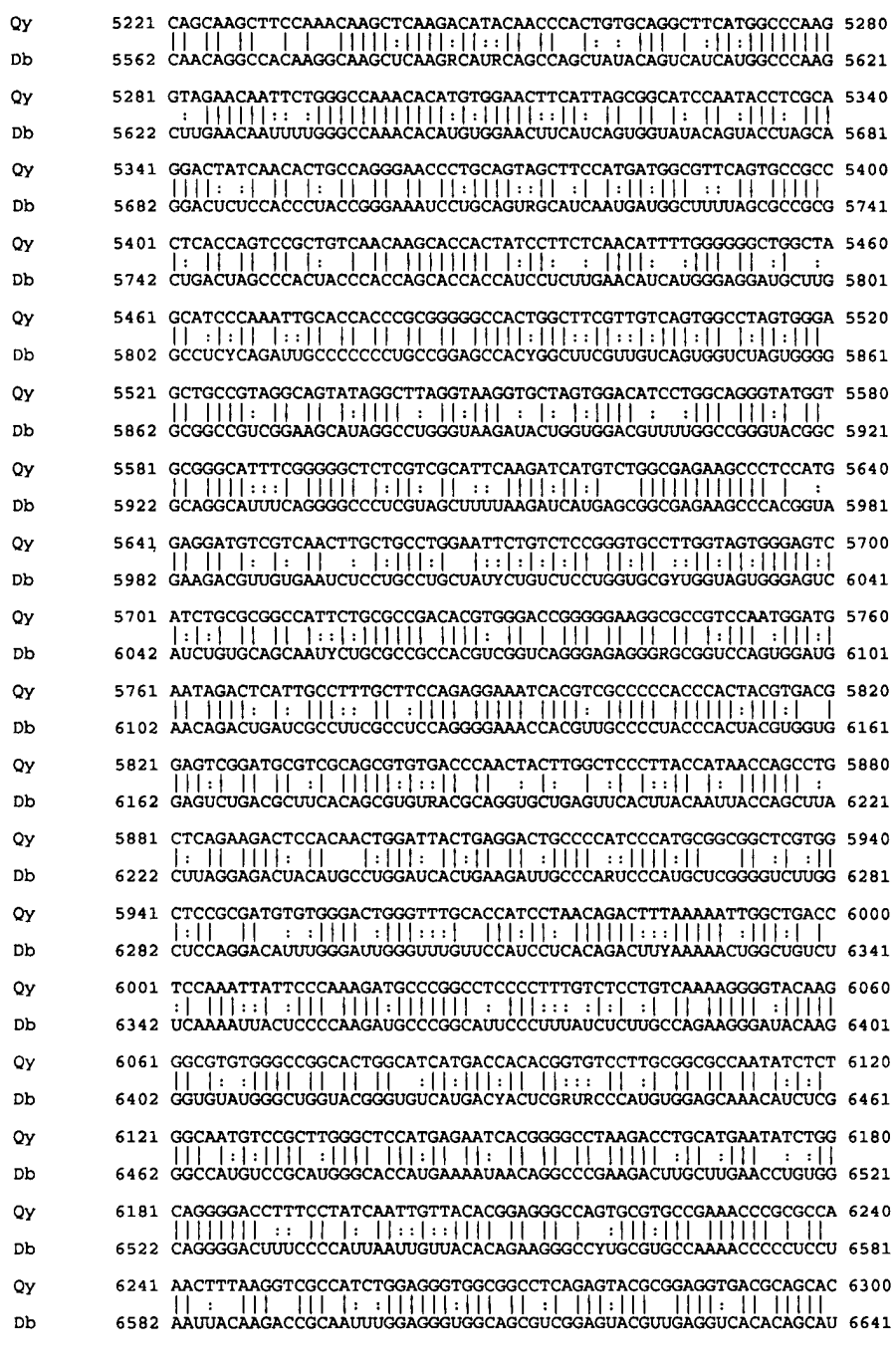
Qy        661 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 720
              |:|||  : || || :|||:||| :| ||||| |          | :  |: :||:|||:|
Db       1002 CUCCAUCUUCCUGGAUGCGUCCCAUGUGAGAAYGAUAAYGGCACCUUGCRUUGCUGGAUA 1061

Qy        721 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 780
              |  |:  |||| :| |:||| |:| | |  | ||| || |:||| |   || :||| ||
Db       1062 CAAGUAACACCCRACGUGGCUGUGAAACACCGCGGUGCGCUCACUCGUAGCCUGCGAACA 1121

Qy        781 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 840
              ||| :|||||:| : |: |:| | || |||  |:||:| ||| : :| |:||| ||  :
Db       1122 CACGUCGACAUGAUCGUAAUGGCAGCUACGGCCUGCUCGGCCUUGUAUGUGGGAGAUGUG 1181
```




Db 3342 CUGCAUGGCCUCCCGGUCUCCGCGAGGCUAGGUAGGGARGUUCUGCUCGGCCCUGCCGAC 3401
Qy 3061 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3120
Db 3402 GGCUACACCUCCAAGGGGUGGAAKCUCCUAGCUCCCAUUACUGCUUACACUCAGCAAACU 3461
Qy 3121 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCT 3180
Db 3462 CGUGGUCUCCUGGGUGCUAUCGUGGUCAGCCUAACGGGCCGCGACAAAAAUGAGCAGGCU 3521
Qy 3181 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3240
Db 3522 GGGCAGGUCCAGGUUCUGUCCUCCGUCACACAAACUUUCUUGGGGACAUCCAUUUCGGGC 3581
Qy 3241 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3300
Db 3582 GUCCUCUGGACAGUAUAUCACGGGGCUGGUAAUAAGACCUUGGCCGGCCCCAAGGGACCA 3641
Qy 3301 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGG 3360
Db 3642 GUCACUCAGAUGUACACCAGCGCAGAAGGGGACCUCGUGGGAUGGCCUAGUCCCCCCGGG 3701
Qy 3361 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3420
Db 3702 ACUAAGUCAUUGGACCCCUGUACCUGCGGGGCCGUAGACCUCUACCUGGUCACCCGAAAC 3761
Qy 3421 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3480
Db 3762 GCUGAUGUCAUUCCGGUCCGGAGGAAAGAUGACCGACGGGGUGCAUUACUCUCGCCAAGG 3821
Qy 3481 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3540
Db 3822 CCCCUCUCAACCCUCAAAGGAUCAUCCGGAGGGCCCGUGCUCUGCUCWAGGGGACACGCC 3881
Qy 3541 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3600
Db 3882 GUGGGCUUGUUCAGAGCGGCCGUGUGUGCCAGGGGUGUAGCCAAAUCUAUUGACUUCAUC 3941
Qy 3601 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 3660
Db 3942 CCCGUCGAAUCACUCGAURUCGCCACACGGACGCCCAGUUUCUCUGACAACAGURCGCCG 4001
Qy 3661 CCTGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 3720
Db 4002 CCAGCUGUGCCCCAGUCUUACCAGGUGGGUUACUUGCACGCACCAACAGGCAGCGGAAAG 4061
Qy 3721 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 3780
Db 4062 AGCACCAAGGUCCCUGCCGCGUAUGCCAGUCAGGGGUAUAAAGUACUCGUACUAAAUCCC 4121
Qy 3781 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 3840
Db 4122 UCUGUCGCGGCCACACUUGGUUUUGGGGCCUACAUGUCCAAAGCCCACGGGAUCAACCCU 4181
Qy 3841 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 3900
Db 4182 AAUAUCAGAACUGGAGUGCGGACCGUUACCACCGGGGACUCUAUCACUUACUCCACUUAU 4241
Qy 3901 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 3960
Db 4242 GGCAAGUUUAUCGCAGAUGGAGGCUGUGCAGCCGGUGCCUAUGACAUCAUCAUAUGCGAC 4301
Qy 3961 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4020
Db 4302 GAAUGCCAUUCAGUGGACGCUACUACCAUCCUUGGCAUUGGAACAGUCCUUGACCAAGCU 4361
Qy 4021 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4080
Db 4362 GAGACCGCAGGCGUCAGGCUAGUGGUYUUGGCCACAGCCACGCCUCCCGGUACGGUGACA 4421
Qy 4081 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4140
Db 4422 ACUCCCCACAGUAACAUAGAGGAGGUGGCCCUUGGUCACGAGGGCGAGAUCCCUUUUUAU 4481
Qy 4141 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4200
Db 4482 GGCAAAGCUAUUCCCCUAGCUUUCAUCAAGGGGGGCAGACACUUGAUCUUUUGCCAUUCA 4541
Qy 4201 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4260
Db 4542 AAGAAGAAGUGCGACGAGCUCGCAGCGGCCCUCCGGGGCAYGGGUGUCAAUGCCGUUGCA 4601
Qy 4261 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4320
Db 4602 UACUAUAGGGGUCUCGACGUCUCCGUUAUACCAACUCAAGGAGACGUGGUGGUUGUCGCC 4661
Qy 4321 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4380
Db 4662 ACUGAUGCCCUAAUGACUGGGUACACCGGCGACUUUGACUCYGUCAUCGACUGUAAUGUU 4721
Qy 4381 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4440
Db 4722 GCAGUCUCUCAGAUUGUUGACUUCAGCCUAGACCCAACCUUCACCAUCACCACUCAAACC 4781
Qy 4441 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4500
Db 4782 GUCCCUCAGGACGCUGUCUCCCGUAGUCAACGUAGAGGGAGAACUGGGAGGGGGCGAUUG 4841
Qy 4501 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4560
Db 4842 GGCRUUUACAGGUAUGUUUCGUCAGGYGARRGGCCGUCUGGGAUGUUCGACAGCGUAGUG 4901
Qy 4561 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4620
Db 4902 CYCUGCGAGUGCUAUGAUGCCGGGGCAGCCUGGUACGAGCUUACACCUGCUGAGACUACG 4961
Qy 4621 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 4680
Db 4962 GUGAGACUCCGGGCYUAUUUCAACACGCCCGGUUUGCCCGUAUGUCAAGACCACCUGGAG 5021
Qy 4681 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 4740
Db 5022 UUCUGGGAAGCGGUCUUUACAGGUCUCACWCACAUURACGCCCACUUCCUCUCCCAGACG 5081
Qy 4741 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 4800
Db 5082 AAGCAAGGAGGAGAAAACUUUGCRUAUCUAACGGCCUACCAGGCCACAGUAUGCGCCAGG 5141
Qy 4801 GCCAAAGCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 4860
Db 5142 GCAAAGGCCCCUCCUCCUUCGUGGGACGUGAUGUGGAAGUGUCUAACUAGGCUCAAACCU 5201
Qy 4861 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 4920
Db 5202 ACACUGACUGGUCCCACCCCCCUCCUGUACCGCUUGGGUGCCGUGACCAAUGAGGUYACC 5261
Qy 4921 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 4980
Db 5262 UUGACGCACCCCGUGACGAAAUACAUCGCCACGUGCAUGCAAGCUGACCUYGAGAUCAUG 5321
Qy 4981 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5040
Db 5322 ACAAGCUCAUGGGUCCUGGCGGGGGGGGUGCUAGCCGCCGUGGCAGCUUACUGCCUGGCG 5381
Qy 5041 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5100
Db 5382 ACUGGCUGCAUUUCCAUCAUUGGCCGCCUACACCUGAAUGAUCGGGUGGUUGUGRCCCCY 5441
Qy 5101 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5160
Db 5442 GACAAGGARAUCUUAUAUGAGGCCUUUGAUGAGAUGGAAGAAUGCGCCUCCAAAGCCGCC 5501
Qy 5161 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5220
Db 5502 CUCAUUGAGGAAGGGCAGCGGAUGGCGGAGAUGCUCAAAUCUAAGAUACAAGGCCUCCUA 5561


Qy 5221 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5280
Db 5562 CAACAGGCCACAAGGCAAGCUCAAGRCAURCAGCCAGCUAUACAGUCAUCAUGGCCCAAG 5621
Qy 5281 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5340
Db 5622 CUUGAACAAUUUUGGGCCAAACACAUGUGGAACUUCAUCAGUGGUAUACAGUACCUAGCA 5681
Qy 5341 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5400
Db 5682 GGACUCUCCACCCUACCGGGAAAUCCUGCAGURGCAUCAAUGAUGGCUUUUAGCGCCGCG 5741
Qy 5401 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5460
Db 5742 CUGACUAGCCCACUACCCACCAGCACCACCAUCCUCUUGAACAUCAUGGGAGGAUGCUUG 5801
Qy 5461 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5520
Db 5802 GCCUCYCAGAUUGCCCCCCCUGCCGGAGCCACYGGCUUCGUUGUCAGUGGUCUAGUGGGG 5861
Qy 5521 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5580
Db 5862 GCGGCCGUCGGAAGCAUAGGCCUGGGUAAGAUACUGGUGGACGUUUUGGCCGGGUACGGC 5921
Qy 5581 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5640
Db 5922 GCAGGCAUUUCAGGGGCCCUCGUAGCUUUUAAGAUCAUGAGCGGCGAGAAGCCCACGGUA 5981
Qy 5641 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 5700
Db 5982 GAAGACGUUGUGAAUCUCCUGCCUGCUAUYCUGUCUCCUGGUGCGYUGGUAGUGGGAGUC 6041
Qy 5701 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 5760
Db 6042 AUCUGUGCAGCAAUYCUGCGCCGCCACGUCGGUCAGGGAGAGGGRGCGGUCCAGUGGAUG 6101
Qy 5761 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 5820
Db 6102 AACAGACUGAUCGCCUUCGCCUCCAGGGGAAACCACGUUGCCCCUACCCACUACGUGGUG 6161
Qy 5821 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 5880
Db 6162 GAGUCUGACGCUUCACAGCGUGURACGCAGGUGCUGAGUUCACUUACAAUUACCAGCUUA 6221
Qy 5881 CTCAGAAGACTCCACAACTGGATTACTGAGGACTGCCCCATCCCATGCGGCGGCTCGTGG 5940
Db 6222 CUUAGGAGACUACAUGCCUGGAUCACUGAAGAUUGCCCARUCCCAUGCUCGGGGUCUUGG 6281
Qy 5941 CTCCGCGATGTGTGGGACTGGGTTTGCACCATCCTAACAGACTTTAAAAATTGGCTGACC 6000
Db 6282 CUCCAGGACAUUUGGGAUUGGGUUUGUUCCAUCCUCACAGACUUYAAAAACUGGCUGUCU 6341
Qy 6001 TCCAAATTATTCCCAAAGATGCCCGGCCTCCCCTTTGTCTCCTGTCAAAAGGGGTACAAG 6060
Db 6342 UCAAAAUUACUCCCCAAGAUGCCCGGCAUUCCCUUUAUCUCUUGCCAGAAGGGAUACAAG 6401
Qy 6061 GGCGTGTGGGCCGGCACTGGCATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6120
Db 6402 GGUGUAUGGGCUGGUACGGGUGUCAUGACYACUCGRURCCCAUGUGGAGCAAACAUCUCG 6461
Qy 6121 GGCAATGTCCGCTTGGGCTCCATGAGAATCACGGGGCCTAAGACCTGCATGAATATCTGG 6180
Db 6462 GGCCAUGUCCGCAUGGGCACCAUGAAAAUAACAGGCCCGAAGACUUGCUUGAACCUGUGG 6521
Qy 6181 CAGGGGACCTTTCCTATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCGCCA 6240
Db 6522 CAGGGGACUUUCCCCAUUAAUUGUUACACAGAAGGGCCYUGCGUGCCAAAACCCCCUCCU 6581
Qy 6241 AACTTTAAGGTCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6300
Db 6582 AAUUACAAGACCGCAAUUUGGAGGGUGGCAGCGUCGGAGUACGUUGAGGUCACACAGCAU 6641
Qy 6301 GGGTCATACCACTACATAACAGGACTCACCACTGATAACTTGAAAGTCCCCTGCCAACTA 6360
Db 6642 GGCUCUUUCUCGUAUGUAACRGGGUUAACCAGUGACAACCUUAAGGUYCCUUGCCAGGUA 6701
Qy 6361 CCCTCTCCCGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCACA 6420
Db 6702 CCAGCUCCAGAAUUUUUCUCUUGGGUGGACGGGGUGCAAAUCCACCGAUUCGCCCCCGUW 6761
Qy 6421 CCGAAGCCGTTTTTCCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6480
Db 6762 CCAGGUCCCUUCUUUCGGGAUGAGGUAACGUUCACCGUAGGCCUUAACUCCUUCGUGGUC 6821
Qy 6481 GGGTCCCAGCTTCCTTGCGACCCTGAACCCGACACAGACGTATTGATGTCCATGCTAACA 6540
Db 6822 GGCUCUCAGCUCCCUUGCGAUCCUGAGCCGGACACCGARGUACUGGCCUCYAUGUUGACA 6881
Qy 6541 GATCCATCTCATATCACGGCGGAGACTGCAGCGCGGCGTTTAGCGCGGGGGTCACCCCCA 6600
Db 6882 GACCCGUCCCACAUCACCGCKGAGGCGGCAGCCAGGCGAUUGGCAAGGGGAUCUCCCCCY 6941
Qy 6601 TCCGAGGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 6660
Db 6942 UCACAGGCUAGCUCCUCAGCGAGCCAGCUCUCUGCCCCGUCCUUGAAGGCUACCUGUACC 7001
Qy 6661 ACCCACGGCAAAGCCTATGATGTGGACATGGTGGATGCTAACCTGTTCATGGGGGGCGAT 6720
Db 7002 ACCCAUAAGACAGCAUAUGAUUGUGACAUGGUGGAUGCYAACCUUUUCAUGGGAGGMGAU 7061
Qy 6721 GTGACTCGGATAGAGTCTGGGTCCAAAGTGGTCGTTCTGGACTCTCTCGACCCAATGGTC 6780
Db 7062 GUGAYCCGGAUUGAGUCUGACUCUAAGGUGAUCGUUCUAGACUCCCUCGAUUCCAUGACU 7121
Qy 6781 GAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCAGAATACATGCTCCCCAAGAAGAGG 6840
Db 7122 GAGGUAGAGGAUGAUCGUGAGCCUUCUGUACCAUCAGAGUACCUGAUCAAGAGGAGAAAG 7181
Qy 6841 TTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCGCTTGTGGAATCG 6900
Db 7182 UUCCCACCGGCGCUGCCUCCUUGGGCCCGUCCAGACUACAAUCCUGUUUUGAUCGAGACA 7241
Qy 6901 TGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTCCCTCCTCCTAGG 6960
Db 7242 UGGAAGAGGCCGGGCUAUGAACCACCCACUGUCCUAGGCUGUGCCCUCCCCCCCCACACYU 7301
Qy 6961 AAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGTGAGGACTCCATA 7020
Db 7302 CAAACGCCAGUGCCUCCACCUCGGAGGCGCCGCGCYAAARUCCUGACCCAGGACRAUGUG 7361
Qy 7021 GGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCCCCAAGCGGCGAT 7080
Db 7362 GAGGGGRUCCUCAGGGAGAUGGCUGACAAAGURCUCAGCCCUCUCCAAGACAACAAUGAC 7421
Qy 7081 TCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACGCCTCCTGATGAG 7140
Db 7422 UCCGGUCACUCCACUGGAGCGGAUACCGGAGGAGACAUCGUCCAGCAACCCUCUGACGAG 7481
Qy 7141 TTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCTCGAGGGGGAGCTTGGA 7200
Db 7482 ACUGCCGCUUCAGAAGCGGGGUCACUGUCCUCCAUGCCUCCCCCUUGAGGGAGAGCCGGGA 7541
Qy 7201 GATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCCAGGGGGGGGTGGCA 7260
Db 7542 GACCCYGACCUGGAGUUUGAACCAGUGGGAUCCGCUCCCCCUUCUGAGGGGGAGUGUGAG 7601
Qy 7261 GCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGACGACTCCGTCGTG 7320
Db 7602 GUCAUUGAUUCGGACUCUAAGUCGUGGUCCACAGUCUCUGAUCAAGAGGAUUCUGUUAUC 7661
Qy 7321 TGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGTAGTCCCGAAGAG 7380
Db 7662 UGCUGCUCUAUGUCAUACUCCUGGACGGGGGCCCUCAUAACACCAUGUGGGCCCGAAGAG 7721
Qy 7381 GAGAAGTTACCGATTAACCCCTTGAGCAACTCCCTGTTGCGATATCACAACAAGGTGTAC 7440

```
Db       7722 GAGAAGUUACCGAUCAACCCUCUGAGUAAUUCGCUCAUGCGGUUCCAUAAYAAGGUGUAC 7781

Qy       7441 TGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTTGATAGGATGCAA 7500

Db       7782 UCCACAACCUCGAGGAGUGCCUCUCUGAGGGCAAAGAAGGUGACUUUUGACAGGGUGCAG 7841

Qy       7501 GTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCGGCCTCCAAGGTC 7560

Db       7842 GUGCUGGACGCACACUAUGACUCAGUCUUGCAGGACGUUAAGCGGGCCGCCUCUAAGGUU 7901

Qy       7561 ACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCCCATTCTGCAAGA 7620

Db       7902 RGUGCGAGGCUCCUCACAGUAGAGGAAGCCUGCGCGCUGACCCCGCCCCACUCCGCCAAA 7961

Qy       7621 TCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGGGCCGTTAACCAC 7680

Db       7962 UCGCGAUACGGAUUUGGGGCAAAAGAGGUGCGCAGCUUAUCCAGGAGGGCCGUUAACCAC 8021

Qy       7681 ATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATTCCCACAACCATT 7740

Db       8022 AUCCGGUCCGUGUGGGAGGACCUCCUGGAAGACCAACRUACCCCAAUUGACACAACUAUC 8081

Qy       7741 ATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAGAAAGCAGCTCGC 7800

Db       8082 AUGGCUAAAAAUGAGGUGUUCUGCAUUGAUCCAACUAARGGUGGGAAAAAGCCAGCUCGC 8141

Qy       7801 CTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCCCTTTATGACATT 7860

Db       8142 CUCAUCGUAUACCCCGACCUUGGGGUCAGGGUGUGCGAAAAGAUGGCCCUCUAUGACAUC 8201

Qy       7861 ACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAGTATTCCCCCGCT 7920

Db       8202 RCACAAAAGCUUCCCAAAGCGAUAAUGGGGCCAUCCUAUGGGUUCCAAUACUCUCCCGCA 8261

Qy       7921 CAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCTATGGGTTTTTCG 7980

Db       8262 GAACGGGUCGAUUUCCUCCUCAAAGCUUGGGGAAGUAAGAAGGACCCAAUGGGGUUCUCG 8321

Qy       7981 TATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGGACTGAGGAGTCC 8040

Db       8322 UAUGACACCCGCUGCUUUGACUCAACCGUCACGGAGAGGGACAUAAGAACAGAAGAAUCC 8381

Qy       8041 ATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACACTCGCTAACTGAG 8100

Db       8382 AUAUAUCAGGCUUGUUCUCUGCCUCAAGAAGCCAGAACUGUCAUACACUCGCUCACUGAG 8441

Qy       8101 AGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGCGGGTACAGGCGT 8160

Db       8442 AGACUUUACGUAGGAGGGCCCAUGACAAACAGCAAAGGGCAAUCCUGCGGCUACAGGCGU 8501

Qy       8161 TGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACATGCTACGTGAAA 8220

Db       8502 UGCCGCGCAAGCGGKGUUUUCACCACCAGCAUGGGGAAUACCAUGACAUGUUACAUCAAA 8561

Qy       8221 GCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTGGTATGCGGCGAT 8280

Db       8562 GCCCUUGCAGCGUGUAAGGCUGCRGGGAUCGUGGACCCUGUUAUGUUGGUGUGUGGAGAC 8621

Qy       8281 GACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGGAACCTGAGAGCC 8340

Db       8622 GACCUGGUCGUCAUCUCAGAGAGCCAAGGUAACGAGGAGGACGAGCGAAACCUGAGAGCU 8681

Qy       8341 TTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCCAGACCGGAGTAT 8400

Db       8682 UUCACGGAGGCUAUGACCAGGUAUUCCGCCCCUCCCGGUGACCUUCCCAGACCGGAAUAU 8741

Qy       8401 GATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTGGGCCCACAAGGC 8460

Db       8742 GACUUGGAGCUUAUAACAUCCUGCUCCUCAAACGUAUCGGUAGCGCUGGACUCUCGGGGU 8801

Qy       8461 CGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGGGCTGCCTGGGAA 8520

Db       8802 CGCCGCCGGUACUUCCUAACCAGAGACCCUACCACUCCAAUCACCCGAGCUGCUUGGGAA 8861

Qy       8521 ACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAGTACGCCCCGACC 8580

Db       8862 ACAGUAAGACACUCCCCUGUCAAUUCUUGGCUGGGCAACAUCAUCCAGUACGCCCCCACA 8921

Qy       8581 ATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATGGCTCAAGACACG 8640

Db       8922 AUCUGGGUCCGGAUGGUCAUAAUGACUCACUUCUUCUCCAUACUAUUGGCCCAGGACACU 8981

Qy       8641 CTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTGAGTCCCTTGGAC 8700

Db       8982 CUGAACCAAAAUCUCAAUUUUGAGAUGUACGGGGCAGUAUACUCGGUCAAUCCAUUAGAC 9041

Qy       8701 CTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTGCACACATACACT 8760

Db       9042 CUACCGGCCAUAAUUGAAAGGCUACAUGGGCUUGAAGCCUUUUCACUGCACACAUACUCU 9101

Qy       8761 CCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCGCCACCCCTCAGA 8820

Db       9102 CCCCACGAACUCUCACGGGUGGCAGCAACUCUCAGAAAACUUGGAGCGCCUCCCCUUAGA 9161

Qy       8821 GCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGTGGGGGGAGAGCG 8880

Db       9162 GCGUGGAAGAGUCGGGCGCGUGCCGUGAGAGCUUCACUCAUCGCCCAAGGAGCGAGGGCG 9221

Qy       8881 GCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 8940

Db       9222 GCCAUUUGUGGCCGCUACCUCUUCAACUGGGCGGUGAAAACAAAGCUCAAACUCACUCCA 9281

Qy       8941 TTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGCGCCGGCGGGGGC 9000

Db       9282 UUGCCCGAGGCGAGCCGCCUGGAUUUAUCCGGGUGGUUCACCGUGGGCGCCGGCGGGGGC 9341

Qy       9001 GACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTTGGCCTACTCCTA 9060

Db       9342 GACAUUUAUCACAGCGUGUCGCAUGCYCGACCCCGCCUAUUACUCCUUUGCCUACUCCUA 9401

Qy       9061 CTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCGG 9099

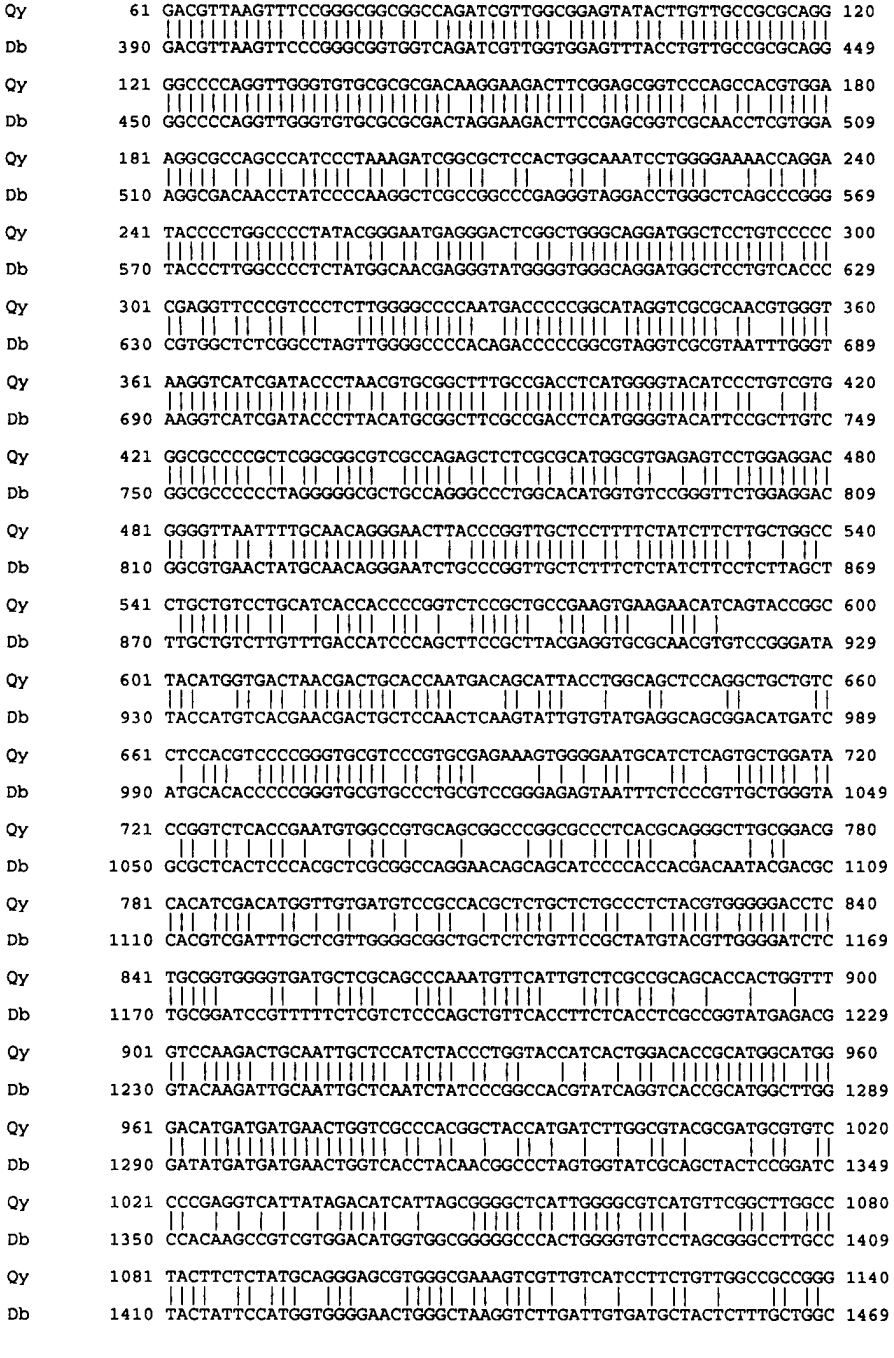
Db       9402 CUUAGCGUAGGAGUAGGCAUCUUUUUACUCCCCGCUCGG 9440

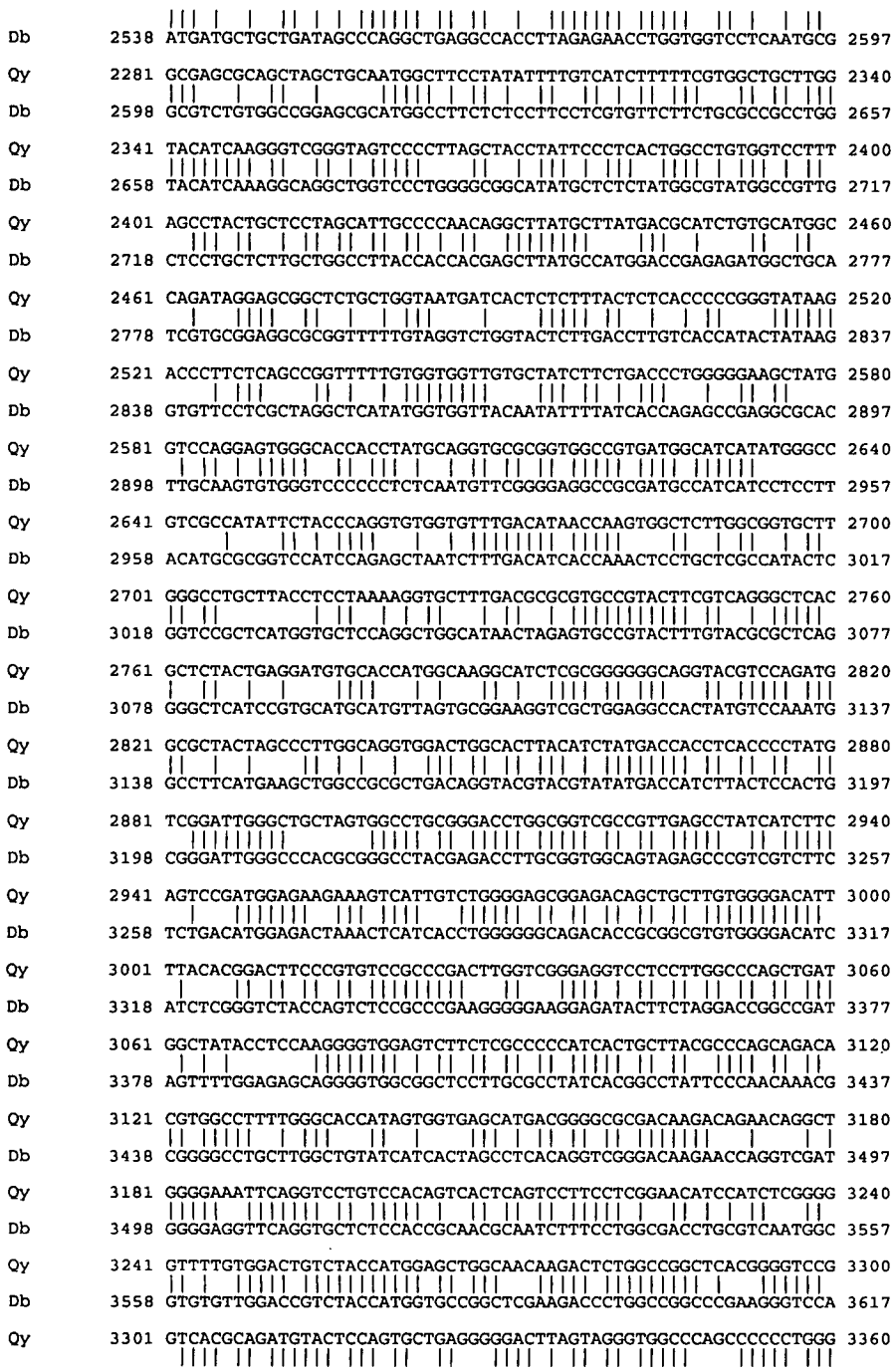

RESULT 5
US-09-827-688-6
; Sequence 6, Application US/09827688
; Patent No. 6821955
; GENERAL INFORMATION:
;  APPLICANT: ORSON, FRANK
;  APPLICANT:  KINSEY, BERMA
;  APPLICANT:  BHOGAL, BALBIR
;  TITLE OF INVENTION: MACROAGGREGATED PROTEIN CONJUGATES AS ORAL GENETIC IMMUNIZATION
;  TITLE OF INVENTION:  AGENTS
;  FILE REFERENCE: P01949US1/10004014
;  CURRENT APPLICATION NUMBER: US/09/827,688
;  CURRENT FILING DATE:  2001-04-06
;  PRIOR APPLICATION NUMBER: 60/195,680
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
;   LENGTH: 9413
;   TYPE: DNA
;   ORGANISM: HEPATITIS C
US-09-827-688-6

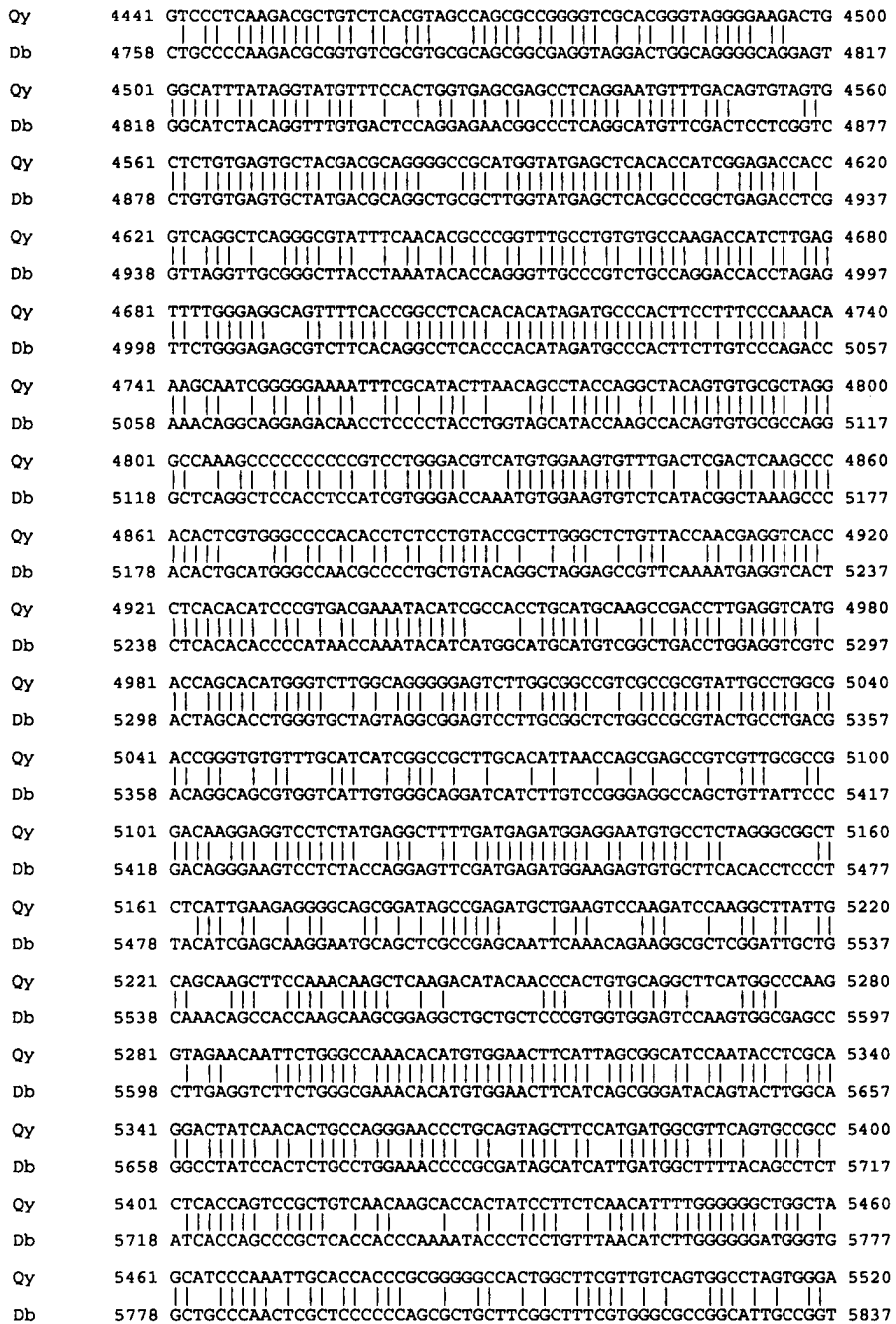
  Query Match            45.5%;  Score 4141.8;  DB 4;  Length 9413;
  Best Local Similarity  66.6%;  Pred. No. 0;
  Matches 6066;  Conservative    0;  Mismatches 2952;  Indels   93;  Gaps    6;

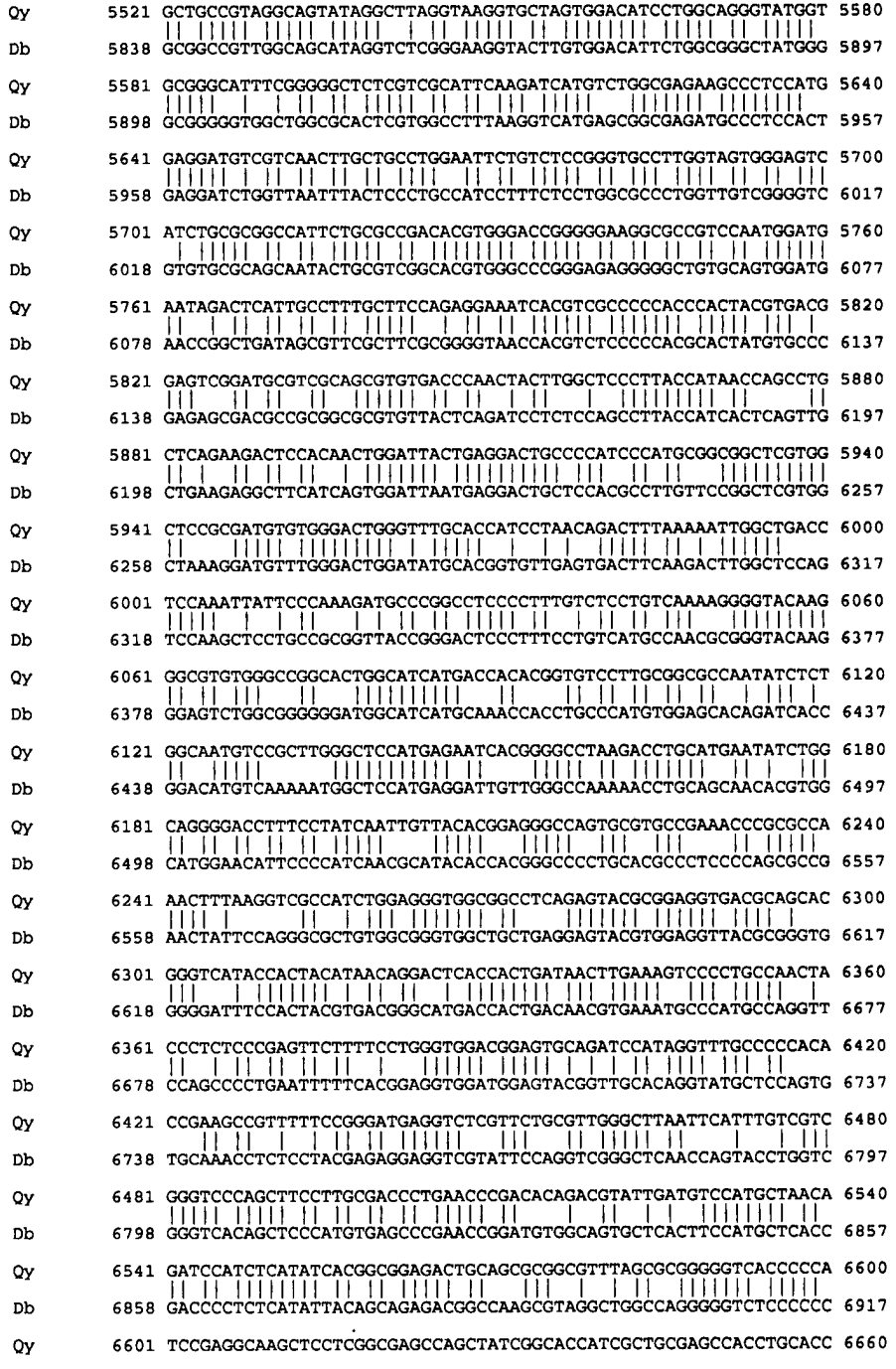
Qy          1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 60

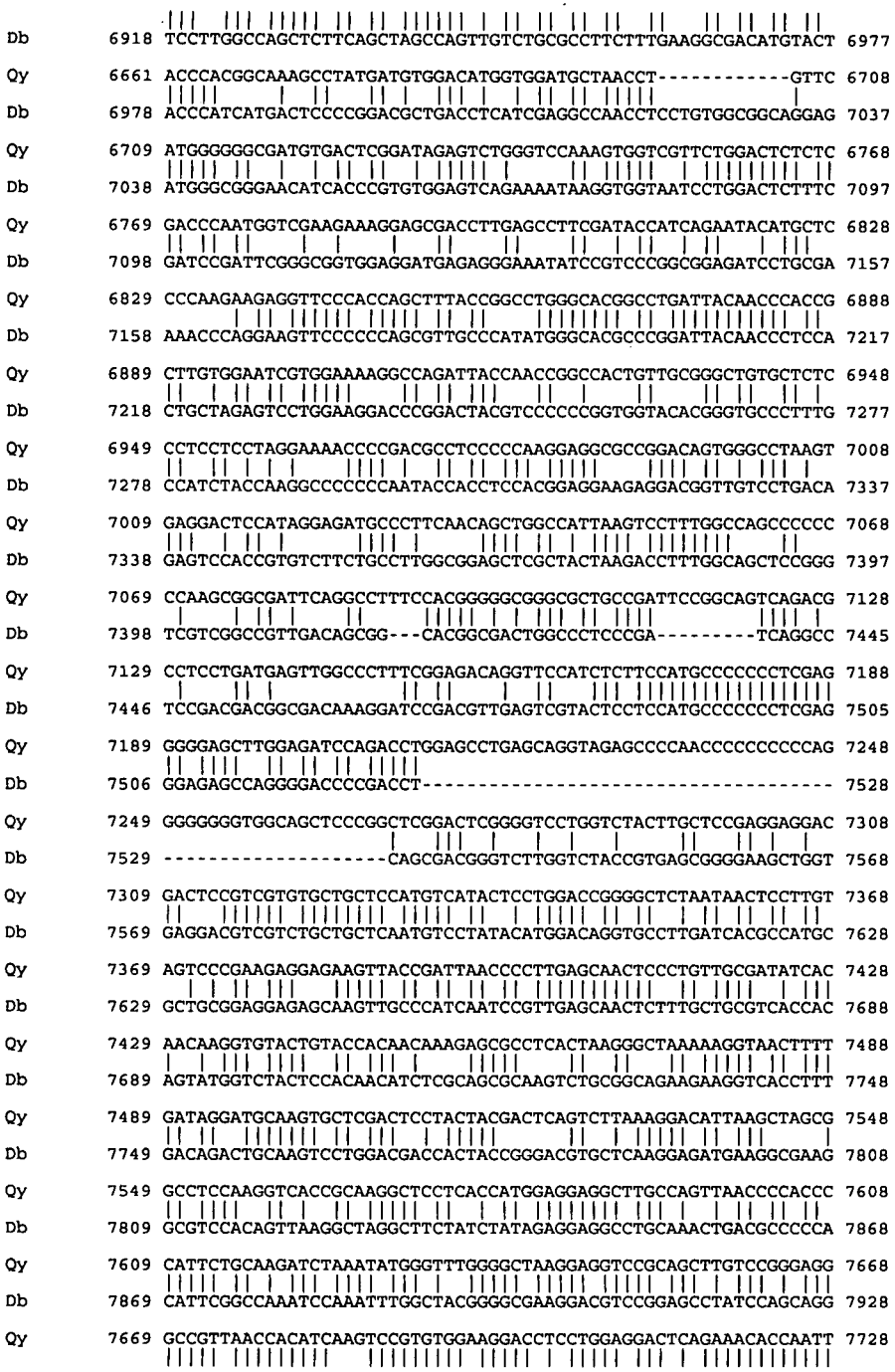
Db        330 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCCGCCCACAG 389
```

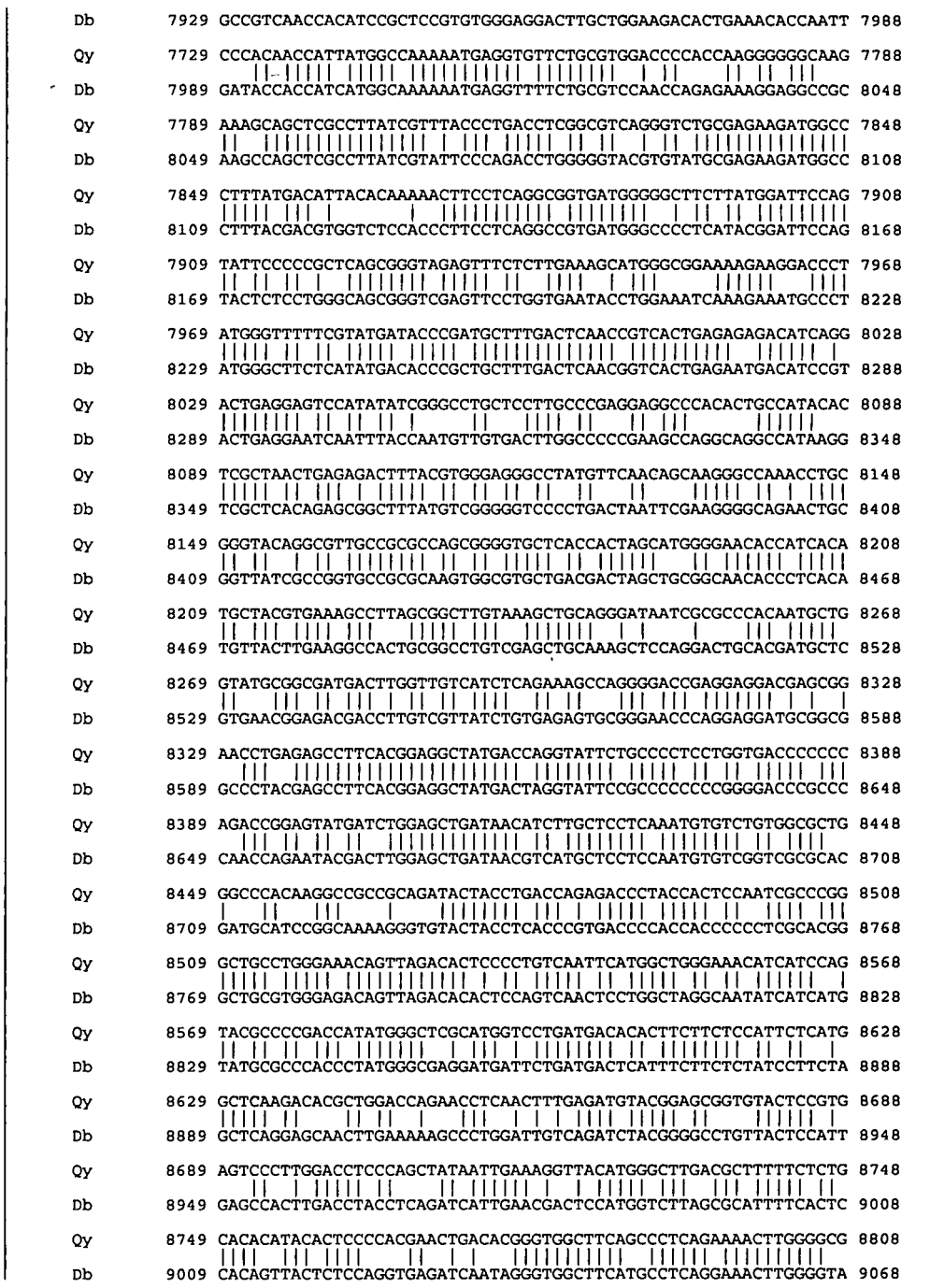

Qy 61 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 120
Db 390 GACGTTAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGG 449
Qy 121 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 180
Db 450 GGCCCCAGGTTGGGTGTGCGCGCGACTAGGAAGACTTCCGAGCGGTCGCAACCTCGTGGA 509
Qy 181 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 240
Db 510 AGGCGACAACCTATCCCCAAGGCTCGCCGGCCCGAGGGTAGGACCTGGGCTCAGCCCGGG 569
Qy 241 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 300
Db 570 TACCCTTGGCCCCTCTATGGCAACGAGGGTATGGGGTGGGCAGGATGGCTCCTGTCACCC 629
Qy 301 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 360
Db 630 CGTGGCTCTCGGCCTAGTTGGGGCCCCACAGACCCCCGGCGTAGGTCGCGTAATTTGGGT 689
Qy 361 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 420
Db 690 AAGGTCATCGATACCCTTACATGCGGCTTCGCCGACCTCATGGGGTACATTCCGCTTGTC 749
Qy 421 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 480
Db 750 GGCGCCCCCCTAGGGGGCGCTGCCAGGGCCCTGGCACATGGTGTCCGGGTTCTGGAGGAC 809
Qy 481 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 540
Db 810 GGCGTGAACTATGCAACAGGGAATCTGCCCGGTTGCTCTTTCTCTATCTTCCTCTTAGCT 869
Qy 541 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 600
Db 870 TTGCTGTCTTGTTTGACCATCCCAGCTTCCGCTTACGAGGTGCGCAACGTGTCCGGGATA 929
Qy 601 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 660
Db 930 TACCATGTCACGAACGACTGCTCCAACTCAAGTATTGTGTATGAGGCAGCGGACATGATC 989
Qy 661 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 720
Db 990 ATGCACACCCCCGGGTGCGTGCCCTGCGTCCGGGAGAGTAATTTCTCCCGTTGCTGGGTA 1049
Qy 721 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 780
Db 1050 GCGCTCACTCCCACGCTCGCGGCCAGGAACAGCAGCATCCCCACCACGACAATACGACGC 1109
Qy 781 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 840
Db 1110 CACGTCGATTTGCTCGTTGGGGCGGCTGCTCTCTGTTCCGCTATGTACGTTGGGGATCTC 1169
Qy 841 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 900
Db 1170 TGCGGATCCGTTTTTCTCGTCTCCCAGCTGTTCACCTTCTCACCTCGCCGGTATGAGACG 1229
Qy 901 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 960
Db 1230 GTACAAGATTGCAATTGCTCAATCTATCCCGGCCACGTATCAGGTCACCGCATGGCTTGG 1289
Qy 961 GACATGATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1020
Db 1290 GATATGATGATGAACTGGTCACCTACAACGGCCCTAGTGGTATCGCAGCTACTCCGGATC 1349
Qy 1021 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1080
Db 1350 CCACAAGCCGTCGTGGACATGGTGGCGGGGGCCCACTGGGGTGTCCTAGCGGGCCTTGCC 1409
Qy 1081 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1140
Db 1410 TACTATTCCATGGTGGGGAACTGGGCTAAGGTCTTGATTGTGATGCTACTCTTTGCTGGC 1469
Qy 1141 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1200
Db 1470 GTTGACGGGCACACCCACGTGACAGGGGGAAGGGTAGCCTCCAGCACCCAGAGCCTCGTG 1529
Qy 1201 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1260
Db 1530 TCCTGGCTCTCACAAGGCCCATCTCAGAAAATCCAACTCGTGAACACCAACGGCAGCTGG 1589
Qy 1261 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1320
Db 1590 CACATCAACAGGACCGCTCTGAATTGCAATGACTCCCTCCAAACTGGGTTCATTGCTGCG 1649
Qy 1321 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1380
Db 1650 CTGTTCTACGCACACAGGTTCAACGCGTCCGGGTGCCCAGAGCGCATGGCTAGCTGCCGC 1709
Qy 1381 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1440
Db 1710 CCCATCGATGAGTTCGCTCAGGGGTGGGGTCCCATCACTCATGATATGCCTGAGAGCTC- 1768
Qy 1441 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1500
Db 1769 -----GGACCAGAGGCCATATTGCTGGCACTACGCGCCTCGACCGTGCGGGATCGTGCCT 1823
Qy 1501 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1560
Db 1824 GCGTCGCAGGTGTGTGGTCCAGTGTATTGCTTCACTCCGAGCCCTGTTGTAGTGGGGACG 1883
Qy 1561 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1620
Db 1884 ACCGATCGTTTCGGCGCTCCTACGTATAGCTGGGGGGAGAATGAGACAGACGTGCTGCTA 1943
Qy 1621 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 1680
Db 1944 CTTAGCAACACGCGGCCGCCTCAAGGCAACTGGTTTGGGTGCACGTGGATGAACAGCACT 2003
Qy 1681 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 1740
Db 2004 GGGTTCACCAAGACGTGCGGGGGCCCTCCGTGCAACATCGGGGGGGTCGGCAACAACA-- 2061
Qy 1741 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 1800
Db 2062 ----CCTTGGTCTGCCCCACGGATTGCTTCCGGAAGCACCCCGAGGCCACTTACACAAAG 2117
Qy 1801 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 1860
Db 2118 TGTGGCTCGGGGCCCTGGTTGACACCCAGGTGCATGGTTGACTACCCATACAGGCTCTGG 2177
Qy 1861 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 1920
Db 2178 CACTACCCCTGCACTGTTAACTTTACCGTCTTTAAGGTCAGGATGTATGTGGGGGGCGTG 2237
Qy 1921 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 1980
Db 2238 GAGCACAGGCTCAATGCTGCATGCAATTGGACTCGAGGAGAGCGCTGTGACTTGGAGGAC 2297
Qy 1981 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2040
Db 2298 AGGGATAGGTCAGAACTCAGCCCGCTGCTGCTGTCTACAACAGAGTGGCAGATACTGCCC 2357
Qy 2041 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2100
Db 2358 TGTTCCTTCACCACCCTACCGGCCCTGTCCACTGGCTTGATCCATCTTCACCGGAACATC 2417
Qy 2101 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2160
Db 2418 GTGGACGTGCAATACCTGTACGGTATAGGGTCGGCAGTTGTCTCCTTTGCAATCAAATGG 2477
Qy 2161 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2220
Db 2478 GAGTATATCCTGTTGCTTTTCCTTCTTCTGGCGGACGCGCGCGTCTGTGCCTGCTTGTGG 2537
Qy 2221 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2280

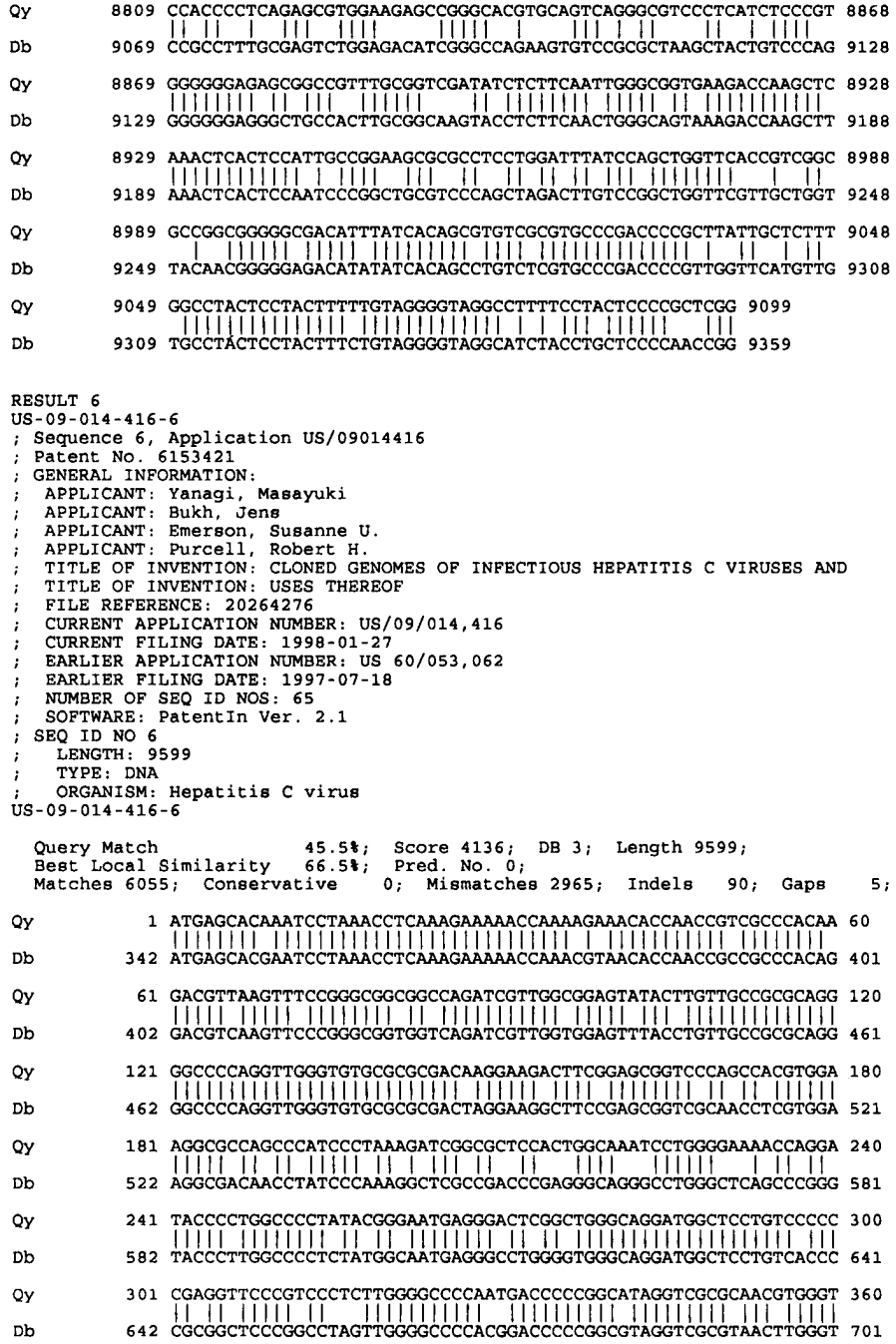

Db 2538 ATGATGCTGCTGATAGCCCAGGCTGAGGCCACCTTAGAGAACCTGGTGGTCCTCAATGCG 2597
Qy 2281 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATATTTTGTCATCTTTTTCGTGGCTGCTTGG 2340
Db 2598 GCGTCTGTGGCCGGAGCGCATGGCCTTCTCTCCTTCCTCGTGTTCTTCTGCGCCGCCTGG 2657
Qy 2341 TACATCAAGGGTCGGGTAGTCCCCTTAGCTACCTATTCCCTCACTGGCCTGTGGTCCTTT 2400
Db 2658 TACATCAAAGGCAGGCTGGTCCCTGGGGCGGCATATGCTCTCTATGGCGTATGGCCGTTG 2717
Qy 2401 AGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTGCATGGC 2460
Db 2718 CTCCTGCTCTTGCTGGCCTTACCACCACGAGCTTATGCCATGGACCGAGAGATGGCTGCA 2777
Qy 2461 CAGATAGGAGCGGCTCTGCTGGTAATGATCACTCTCTTTACTCTCACCCCCGGGTATAAG 2520
Db 2778 TCGTGCGGAGGCGCGGTTTTTGTAGGTCTGGTACTCTTGACCTTGTCACCATACTATAAG 2837
Qy 2521 ACCCTTCTCAGCCGGTTTTTGTGGTGGTTGTGCTATCTTCTGACCCTGGGGGAAGCTATG 2580
Db 2838 GTGTTCCTCGCTAGGCTCATATGGTGGTTACAATATTTTATCACCAGAGCCGAGGCGCAC 2897
Qy 2581 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGCATCATATGGGCC 2640
Db 2898 TTGCAAGTGTGGGTCCCCCCTCTCAATGTTCGGGGAGGCCGCGATGCCATCATCCTCCTT 2957
Qy 2641 GTCGCCATATTCTACCCAGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 2700
Db 2958 ACATGCGCGGTCCATCCAGAGCTAATCTTTGACATCACCAAACTCCTGCTCGCCATACTC 3017
Qy 2701 GGGCCTGCTTACCTCCTAAAAGGTGCTTTGACGCGCGTGCCGTACTTCGTCAGGGCTCAC 2760
Db 3018 GGTCCGCTCATGGTGCTCCAGGCTGGCATAACTAGAGTGCCGTACTTTGTACGCGCTCAG 3077
Qy 2761 GCTCTACTGAGGATGTGCACCATGGCAAGGCATCTCGCGGGGGGCAGGTACGTCCAGATG 2820
Db 3078 GGGCTCATCCGTGCATGCATGTTAGTGCGGAAGGTCGCTGGAGGCCACTATGTCCAAATG 3137
Qy 2821 GCGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 2880
Db 3138 GCCTTCATGAAGCTGGCCGCGCTGACAGGTACGTACGTATATGACCATCTTACTCCACTG 3197
Qy 2881 TCGGATTGGGCTGCTAGTGGCCTGCGGGACCTGGCGGTCGCCGTTGAGCCTATCATCTTC 2940
Db 3198 CGGGATTGGGCCCACGCGGGCCTACGAGACCTTGCGGTGGCAGTAGAGCCCGTCGTCTTC 3257
Qy 2941 AGTCCGATGGAGAAGAAAGTCATTGTCTGGGGAGCGGAGACAGCTGCTTGTGGGGACATT 3000
Db 3258 TCTGACATGGAGACTAAACTCATCACCTGGGGGGCAGACACCGCGGCGTGTGGGGACATC 3317
Qy 3001 TTACACGGACTTCCCGTGTCCGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3060
Db 3318 ATCTCGGGTCTACCAGTCTCCGCCCGAAGGGGGAAGGAGATACTTCTAGGACCGGCCGAT 3377
Qy 3061 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3120
Db 3378 AGTTTTGGAGAGCAGGGGTGGCGGCTCCTTGCGCCTATCACGGCCTATTCCCAACAAACG 3437
Qy 3121 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCT 3180
Db 3438 CGGGGCCTGCTTGGCTGTATCATCACTAGCCTCACAGGTCGGGACAAGAACCAGGTCGAT 3497
Qy 3181 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3240
Db 3498 GGGGAGGTTCAGGTGCTCTCCACCGCAACGCAATCTTTCCTGGCGACCTGCGTCAATGGC 3557
Qy 3241 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3300
Db 3558 GTGTGTTGGACCGTCTACCATGGTGCCGGCTCGAAGACCCTGGCCGGCCCGAAGGGTCCA 3617
Qy 3301 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGG 3360
Db 3618 ATCACCCAAATGTACACCAATGTAGACCAGGACCTCGTCGGCTGGCCGGCGCCCCCCGGG 3677
Qy 3361 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3420
Db 3678 GCGCGCTCCATGACACCGTGCACCTGCGGCAGCTCGGACCTTTACTTGGTCACGAGGCAT 3737
Qy 3421 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3480
Db 3738 GCTGATGTCGTTCCGGTGCGCCGGCGGGGCGACAGCAGGGGGAGCCTGCTTTCCCCCAGG 3797
Qy 3481 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3540
Db 3798 CCCATCTCCTACCTGAAGGGCTCCTCGGGTGGACCACTGCTTTGCCCTTCGGGGCACGTT 3857
Qy 3541 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3600
Db 3858 GTAGGCATCTTCCGGGCTGCTGTGTGCACCCGGGGGGTTGCGAAGGCGGTGGACTTCATA 3917
Qy 3601 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 3660
Db 3918 CCCGTTGAGTCTATGGAAACTACCATGCGGTCTCCGGTCTTCACAGACAACTCATCCCCT 3977
Qy 3661 CCTGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 3720
Db 3978 CCGGCCGTACCGCAAACATTCCAAGTGGCACATTTACACGCTCCCACTGGCAGCGGCAAG 4037
Qy 3721 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 3780
Db 4038 AGCACCAAAGTGCCGGCTGCATATGCAGCCCAAGGGTACAAGGTGCTCGTCCTAAACCCG 4097
Qy 3781 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 3840
Db 4098 TCCGTTGCCGCCACATTGGGCTTTGGAGCGTATATGTCCAAGGCACATGGCATCGAGCCT 4157
Qy 3841 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 3900
Db 4158 AACATCAGAACTGGGGTAAGGACCATCACCACGGGCGGCCCCATCACGTACTCCACCTAT 4217
Qy 3901 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 3960
Db 4218 TGCAAGTTCCTTGCCGACGGTGGATGCTCCGGGGGCGCCTATGACATCATAATATGTGAT 4277
Qy 3961 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4020
Db 4278 GAATGCCACTCAACTGACTCGACTACCATCTTGGGCATCGGCACAGTCCTGGATCAGGCA 4337
Qy 4021 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4080
Db 4338 GAGACGGCTGGAGCGCGGCTCGTCGTGCTCGCCACCGCCACGCCTCCGGGATCGATCACC 4397
Qy 4081 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4140
Db 4398 GTGCCACACCCCAACATCGAGGAAGTGGCCCTGTCCAACACTGGAGAGATTCCCTTCTAT 4457
Qy 4141 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4200
Db 4458 GGCAAAGCCATCCCCATTGAGGCCATCAAGGGGGGAAGGCATCTCATCTTCTGCCATTCC 4517
Qy 4201 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4260
Db 4518 AAGAAGAAGTGTGACGAGCTCGCCGCAAAGCTGACAGGCCTCGGACTCAATGCTGTAGCG 4577
Qy 4261 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4320
Db 4578 TATTACCGGGGTCTCGATGTGTCCGTCATACCGACTAGCGGAGACGTCGTTGTCGTGGCA 4637
Qy 4321 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4380
Db 4638 ACAGACGCTCTAATGACGGGTTTTACCGGCGACTTTGACTCAGTGATCGACTGCAACACA 4697
Qy 4381 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4440
Db 4698 TGTGTCACCCAGACAGTCGATTTCAGCTTGGATCCCACCTTCACCATTGAGACGACAACG 4757





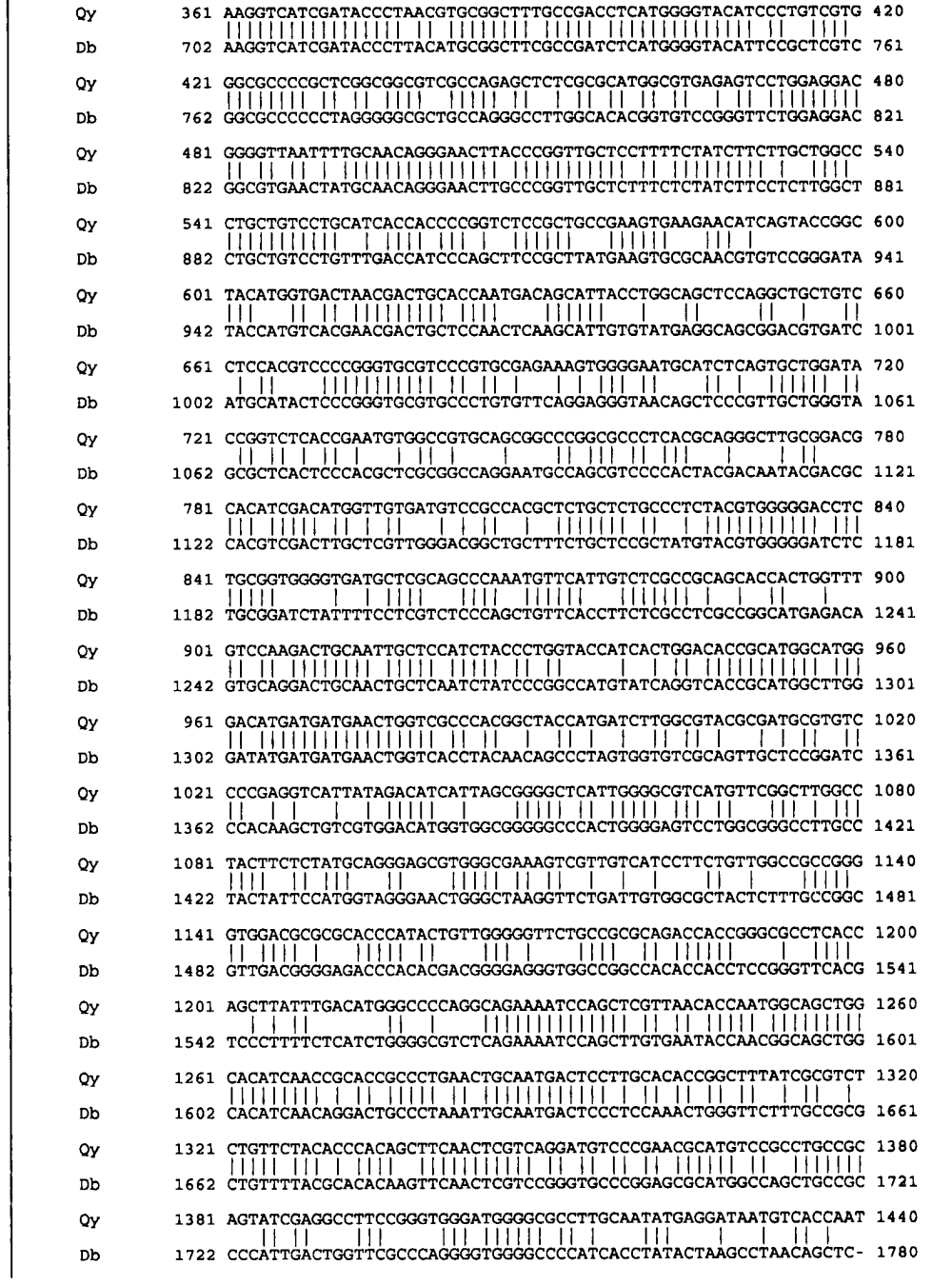


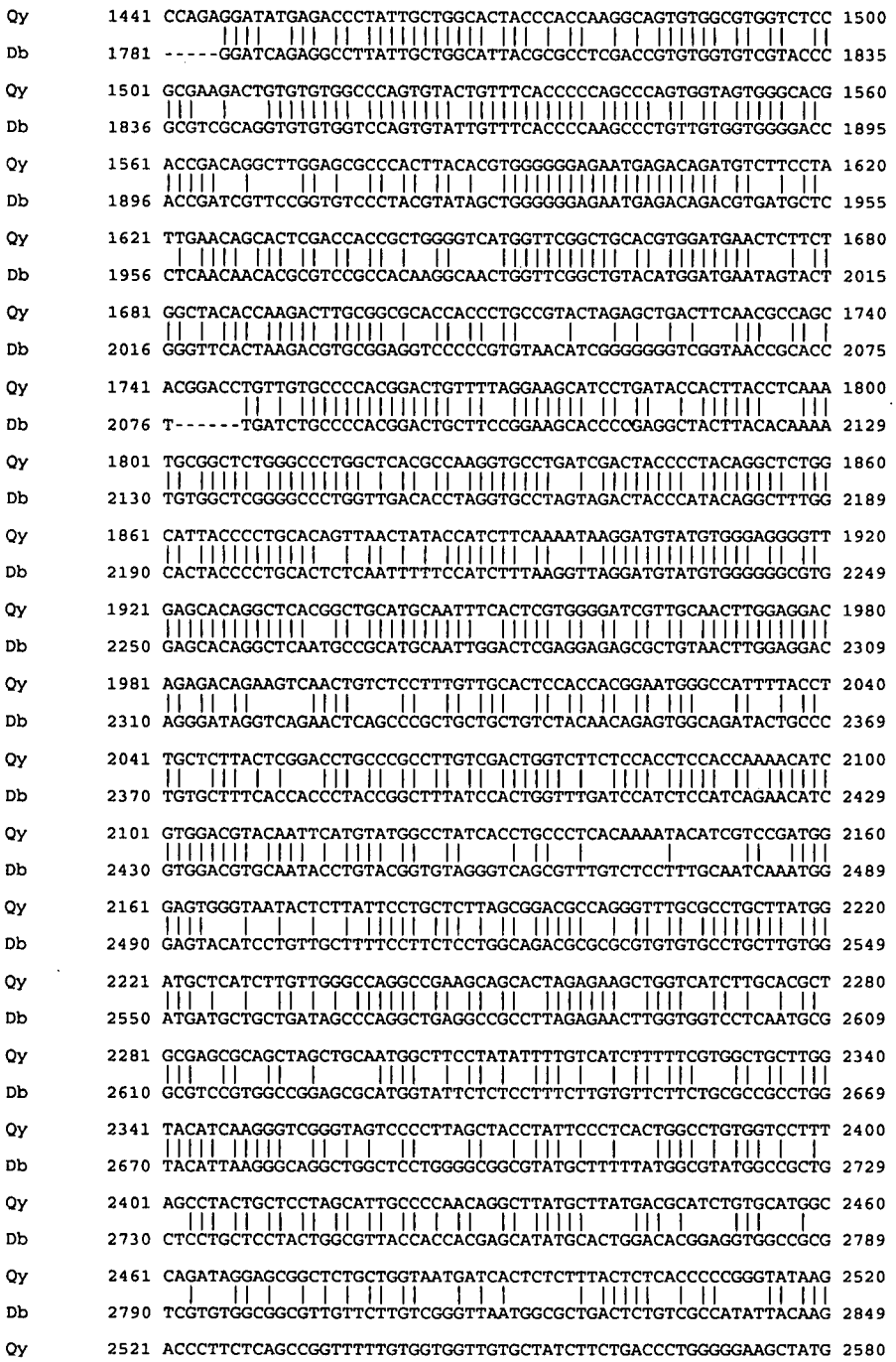


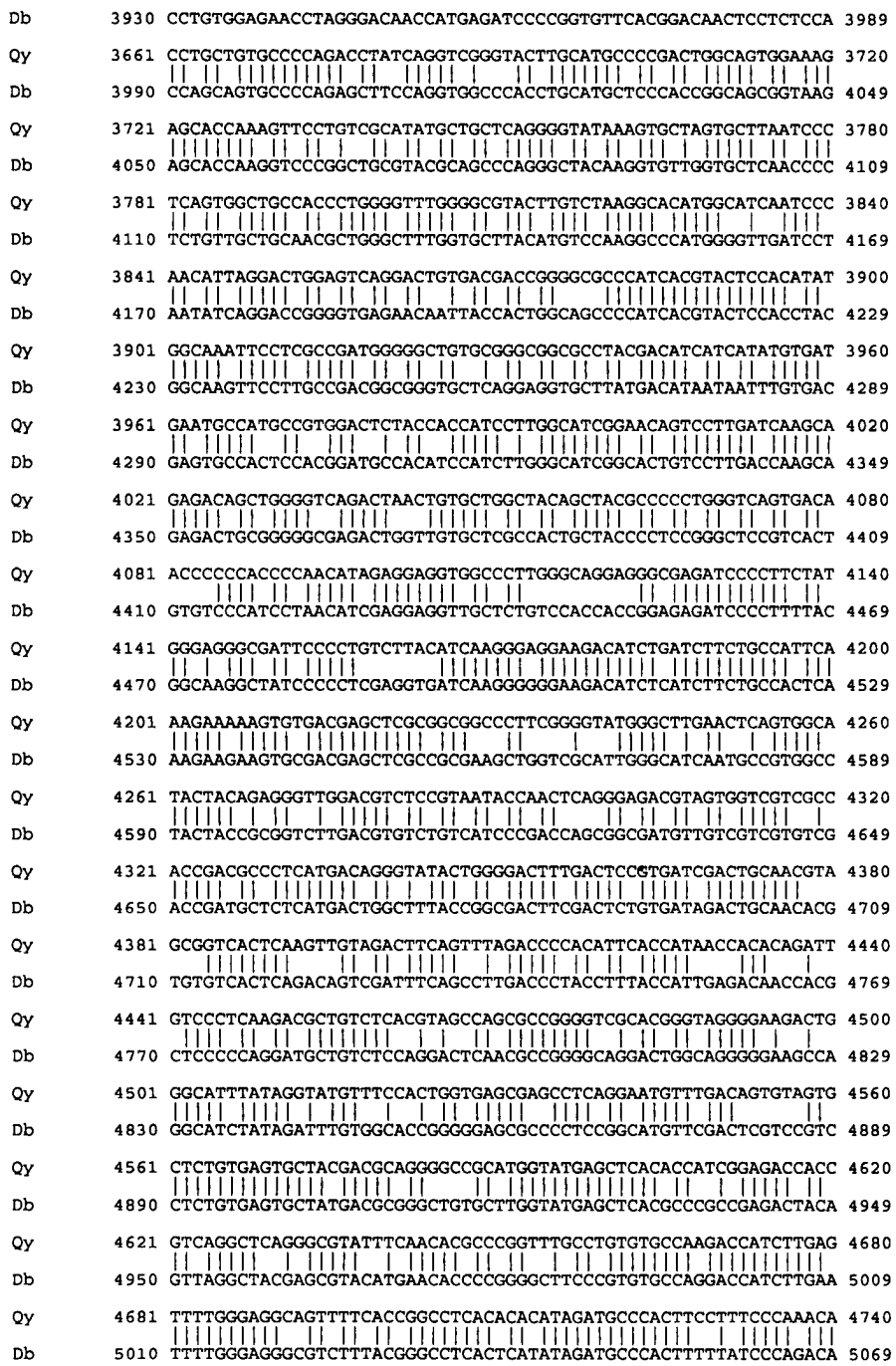


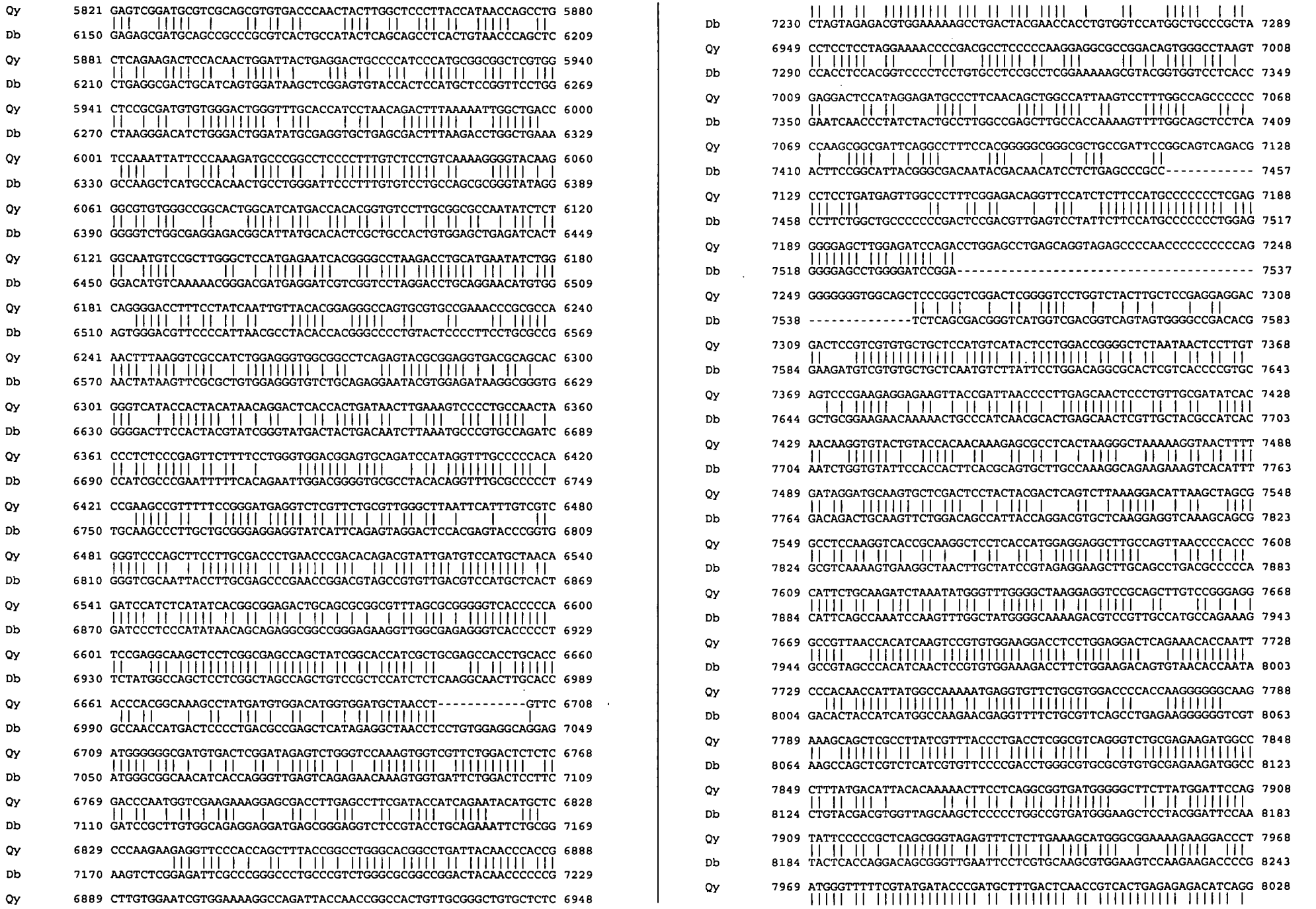
RESULT 6
US-09-014-416-6
; Sequence 6, Application US/09014416
; Patent No. 6153421
; GENERAL INFORMATION:
; APPLICANT: Yanagi, Masayuki
; APPLICANT: Bukh, Jens
; APPLICANT: Emerson, Susanne U.
; APPLICANT: Purcell, Robert H.
; TITLE OF INVENTION: CLONED GENOMES OF INFECTIOUS HEPATITIS C VIRUSES AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: 20264276
; CURRENT APPLICATION NUMBER: US/09/014,416
; CURRENT FILING DATE: 1998-01-27
; EARLIER APPLICATION NUMBER: US 60/053,062
; EARLIER FILING DATE: 1997-07-18
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 9599
; TYPE: DNA
; ORGANISM: Hepatitis C virus
US-09-014-416-6

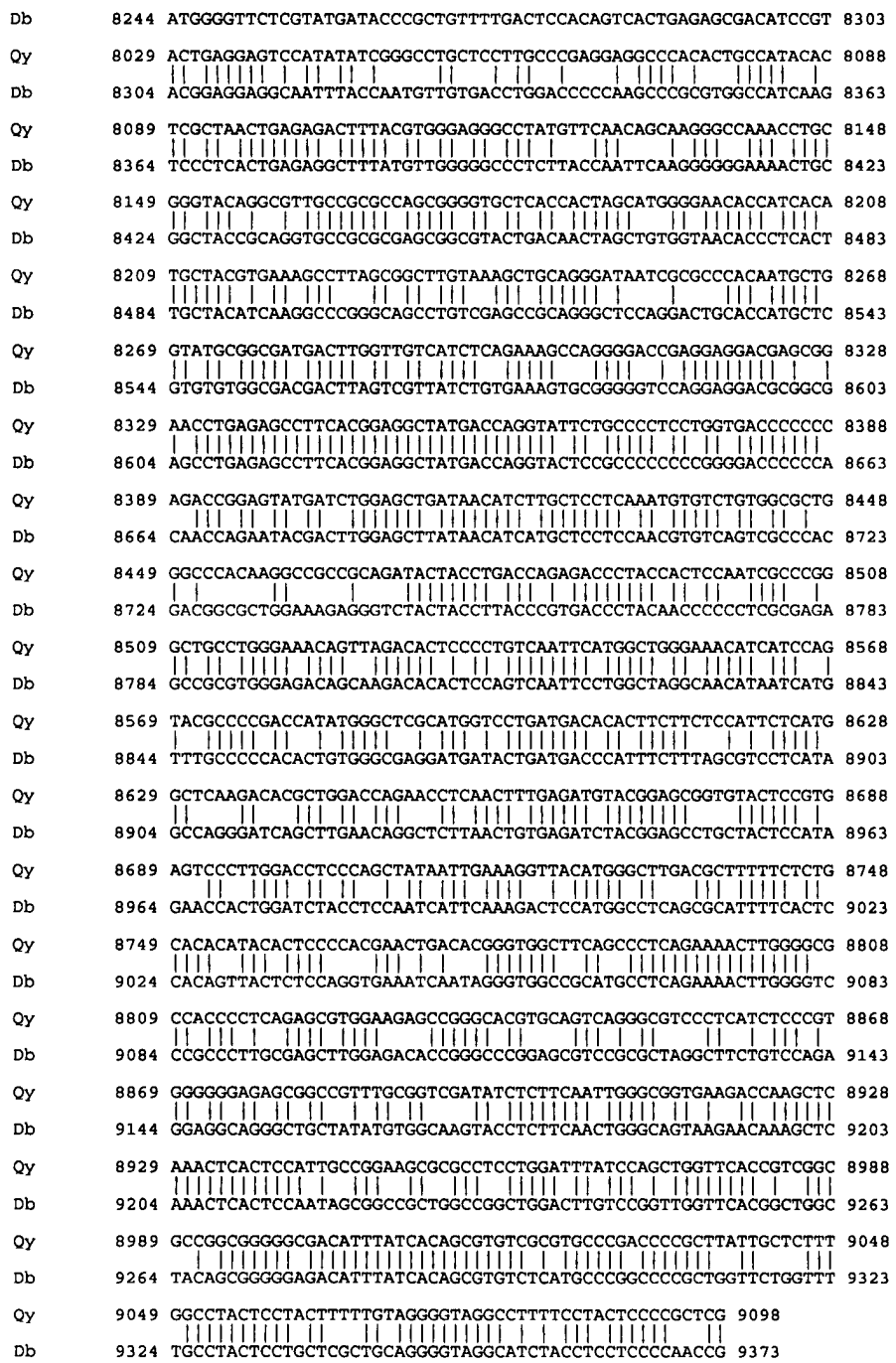
Query Match 45.5%; Score 4136; DB 3; Length 9599;
Best Local Similarity 66.5%; Pred. No. 0;
Matches 6055; Conservative 0; Mismatches 2965; Indels 90; Gaps 5;




Qy 1441 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1500
Db 1781 -----GGATCAGAGGCCTTATTGCTGGCATTACGCGCCTCGACCGTGTGGTGTCGTACCC 1835
Qy 1501 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1560
Db 1836 GCGTCGCAGGTGTGTGGTCCAGTGTATTGTTTCACCCCAAGCCCTGTTGTGGTGGGGACC 1895
Qy 1561 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1620
Db 1896 ACCGATCGTTCCGGTGTCCCTACGTATAGCTGGGGGGAGAATGAGACAGACGTGATGCTC 1955
Qy 1621 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 1680
Db 1956 CTCAACAACACGCGTCCGCCACAAGGCAACTGGTTCGGCTGTACATGGATGAATAGTACT 2015
Qy 1681 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 1740
Db 2016 GGGTTCACTAAGACGTGCGGAGGTCCCCCGTGTAACATCGGGGGGGTCGGTAACCGCACC 2075
Qy 1741 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 1800
Db 2076 T------TGATCTGCCCCACGGACTGCTTCCGGAAGCACCCCGAGGCTACTTACACAAAA 2129
Qy 1801 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 1860
Db 2130 TGTGGCTCGGGGCCCTGGTTGACACCTAGGTGCCTAGTAGACTACCCATACAGGCTTTGG 2189
Qy 1861 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 1920
Db 2190 CACTACCCCTGCACTCTCAATTTTTCCATCTTTAAGGTTAGGATGTATGTGGGGGGCGTG 2249
Qy 1921 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 1980
Db 2250 GAGCACAGGCTCAATGCCGCATGCAATTGGACTCGAGGAGAGCGCTGTAACTTGGAGGAC 2309
Qy 1981 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2040
Db 2310 AGGGATAGGTCAGAACTCAGCCCGCTGCTGCTGTCTACAACAGAGTGGCAGATACTGCCC 2369
Qy 2041 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2100
Db 2370 TGTGCTTTCACCACCCTACCGGCTTTATCCACTGGTTTGATCCATCTCCATCAGAACATC 2429
Qy 2101 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2160
Db 2430 GTGGACGTGCAATACCTGTACGGTGTAGGGTCAGCGTTTGTCTCCTTTGCAATCAAATGG 2489
Qy 2161 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2220
Db 2490 GAGTACATCCTGTTGCTTTTCCTTCTCCTGGCAGACGCGCGCGTGTGTGCCTGCTTGTGG 2549
Qy 2221 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2280
Db 2550 ATGATGCTGCTGATAGCCCAGGCTGAGGCCGCCTTAGAGAACTTGGTGGTCCTCAATGCG 2609
Qy 2281 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATATTTTGTCATCTTTTTCGTGGCTGCTTGG 2340
Db 2610 GCGTCCGTGGCCGGAGCGCATGGTATTCTCTCCTTTCTTGTGTTCTTCTGCGCCGCCTGG 2669
Qy 2341 TACATCAAGGGTCGGGTAGTCCCCTTAGCTACCTATTCCCTCACTGGCCTGTGGTCCTTT 2400
Db 2670 TACATTAAGGGCAGGCTGGCTCCTGGGGCGGCGTATGCTTTTTATGGCGTATGGCCGCTG 2729
Qy 2401 AGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTGCATGGC 2460
Db 2730 CTCCTGCTCCTACTGGCGTTACCACCACGAGCATATGCACTGGACACGGAGGTGGCCGCG 2789
Qy 2461 CAGATAGGAGCGGCTCTGCTGGTAATGATCACTCTCTTTACTCTCACCCCCGGGTATAAG 2520
Db 2790 TCGTGTGGCGGCGTTGTTCTTGTCGGGTTAATGGCGCTGACTCTGTCGCCATATTACAAG 2849
Qy 2521 ACCCTTCTCAGCCGGTTTTTGTGGTGGTTGTGCTATCTTCTGACCCTGGGGGAAGCTATG 2580
Db 2850 CGCTATATCAGCTGGTGCATGTGGTGGCTTCAGTATTTTCTGACCAGAGTAGAAGCGCAA 2909
Qy 2581 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGCATCATATGGGCC 2640
Db 2910 CTGCACGTGTGGGTTCCCCCCCTCAACGTCCGGGGGGGGCGCGATGCCGTCATCTTACTC 2969
Qy 2641 GTCGCCATATTCTACCCAGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 2700
Db 2970 ATGTGTGTAGTACACCCGACCCTGGTATTTGACATCACCAAACTACTCCTGGCCATCTTC 3029
Qy 2701 GGGCCTGCTTACCTCCTAAAAGGTGCTTTGACGCGCGTGCCGTACTTCGTCAGGGCTCAC 2760
Db 3030 GGACCCCTTTGGATTCTTCAAGCCAGTTTGCTTAAAGTCCCCTACTTCGTGCGCGTTCAA 3089
Qy 2761 GCTCTACTGAGGATGTGCACCATGGCAAGGCATCTCGCGGGGGGCAGGTACGTCCAGATG 2820
Db 3090 GGCCTTCTCCGGATCTGCGCGCTAGCGCGGAAGATAGCCGGAGGTCATTACGTGCAAATG 3149
Qy 2821 GCGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 2880
Db 3150 GCCATCATCAAGTTAGGGGCGCTTACTGGCACCTATGTGTATAACCATCTCACCCCTCTT 3209
Qy 2881 TCGGATTGGGCTGCTAGTGGCCTGCGGGACCTGGCGGTCGCCGTTGAGCCTATCATCTTC 2940
Db 3210 CGAGACTGGGCGCACAACGGCCTGCGAGATCTGGCCGTGGCTGTGGAACCAGTCGTCTTC 3269
Qy 2941 AGTCCGATGGAGAAGAAAGTCATTGTCTGGGGAGCGGAGACAGCTGCTTGTGGGGACATT 3000
Db 3270 TCCCGAATGGAGACCAAGCTCATCACGTGGGGGGCAGATACCGCCGCGTGCGGTGACATC 3329
Qy 3001 TTACACGGACTTCCCGTGTCCGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3060
Db 3330 ATCAACGGCTTGCCCGTCTCTGCCCGTAGGGGCCAGGAGATACTGCTTGGGCCAGCCGAC 3389
Qy 3061 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3120
Db 3390 GGAATGGTCTCCAAGGGGTGGAGGTTGCTGGCGCCCATCACGGCGTACGCCCAGCAGACG 3449
Qy 3121 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCT 3180
Db 3450 AGAGGCCTCCTAGGGTGTATAATCACCAGCCTGACTGGCCGGGACAAAAACCAAGTGGAG 3509
Qy 3181 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3240
Db 3510 GGTGAGGTCCAGATCGTGTCAACTGCTACCCAAACCTTCCTGGCAACGTGCATCAATGGG 3569
Qy 3241 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3300
Db 3570 GTATGCTGGACTGTCTACCACGGGGCCGGAACGAGGACCATCGCATCACCCAAGGGTCCT 3629
Qy 3301 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGG 3360
Db 3630 GTCATCCAGATGTATACCAATGTGGACCAAGACCTTGTGGGCTGGCCCGCTCCTCAAGGT 3689
Qy 3361 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3420
Db 3690 TCCCGCTCATTGACACCCTGTACCTGCGGCTCCTCGGACCTTTACCTGGTCACGAGGCAC 3749
Qy 3421 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3480
Db 3750 GCCGATGTCATTCCCGTGCGCCGGCGAGGTGATAGCAGGGGTAGCCTGCTTTCGCCCCGG 3809
Qy 3481 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3540
Db 3810 CCCATTTCCTACTTGAAAGGCTCCTCGGGGGGTCCGCTGTTGTGCCCCGCGGGACACGCC 3869
Qy 3541 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3600
Db 3870 GTGGGCCTATTCAGGGCCGCGGTGTGCACCCGTGGAGTGGCTAAAGCGGTGGACTTTATC 3929
Qy 3601 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 3660

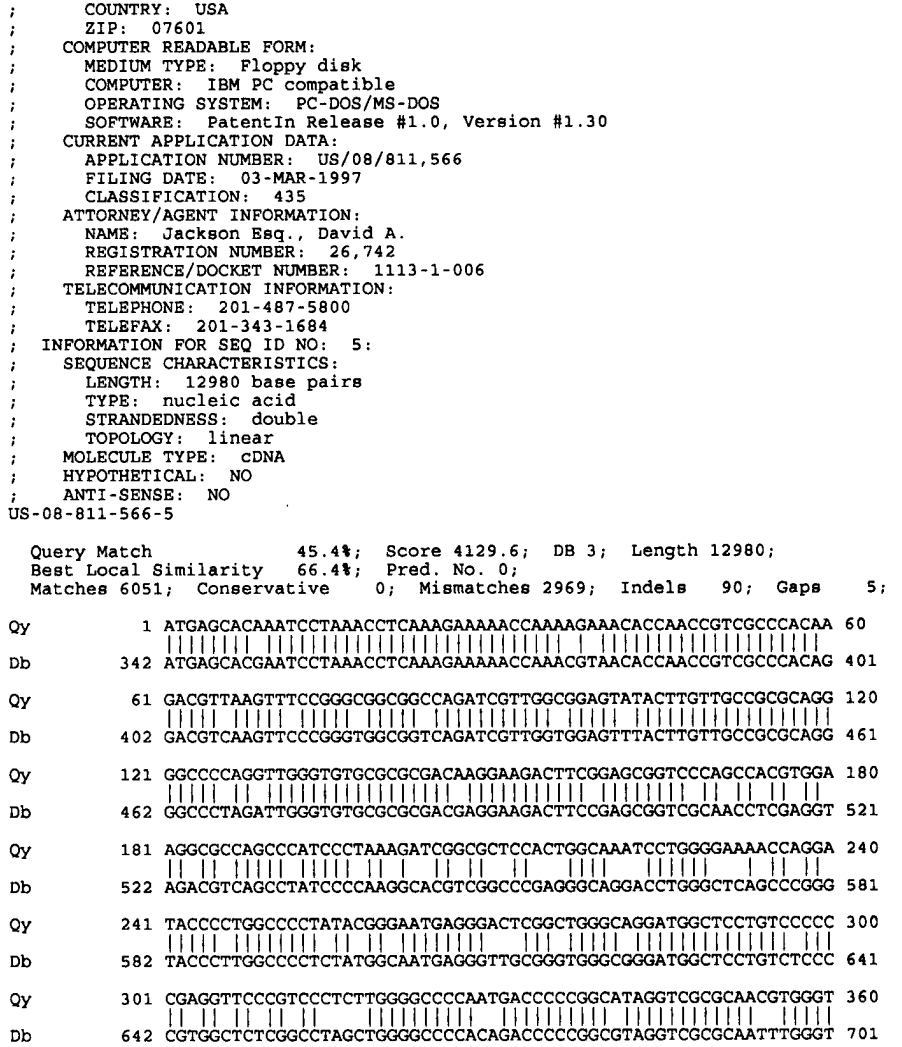

Db 3930 CCTGTGGAGAACCTAGGGACAACCATGAGATCCCCGGTGTTCACGGACAACTCCTCTCCA 3989
Qy 3661 CCTGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 3720
Db 3990 CCAGCAGTGCCCCAGAGCTTCCAGGTGGCCCACCTGCATGCTCCCACCGGCAGCGGTAAG 4049
Qy 3721 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 3780
Db 4050 AGCACCAAGGTCCCGGCTGCGTACGCAGCCCAGGGCTACAAGGTGTTGGTGCTCAACCCC 4109
Qy 3781 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 3840
Db 4110 TCTGTTGCTGCAACGCTGGGCTTTGGTGCTTACATGTCCAAGGCCCATGGGGTTGATCCT 4169
Qy 3841 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 3900
Db 4170 AATATCAGGACCGGGGTGAGAACAATTACCACTGGCAGCCCCATCACGTACTCCACCTAC 4229
Qy 3901 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 3960
Db 4230 GGCAAGTTCCTTGCCGACGGCGGGTGCTCAGGAGGTGCTTATGACATAATAATTTGTGAC 4289
Qy 3961 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4020
Db 4290 GAGTGCCACTCCACGGATGCCACATCCATCTTGGGCATCGGCACTGTCCTTGACCAAGCA 4349
Qy 4021 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4080
Db 4350 GAGACTGCGGGGGCGAGACTGGTTGTGCTCGCCACTGCTACCCCTCCGGGCTCCGTCACT 4409
Qy 4081 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4140
Db 4410 GTGTCCCATCCTAACATCGAGGAGGTTGCTCTGTCCACCACCGGAGAGATCCCCTTTTAC 4469
Qy 4141 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4200
Db 4470 GGCAAGGCTATCCCCCTCGAGGTGATCAAGGGGGGAAGACATCTCATCTTCTGCCACTCA 4529
Qy 4201 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4260
Db 4530 AAGAAGAAGTGCGACGAGCTCGCCGCGAAGCTGGTCGCATTGGGCATCAATGCCGTGGCC 4589
Qy 4261 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4320
Db 4590 TACTACCGCGGTCTTGACGTGTCTGTCATCCCGACCAGCGGCGATGTTGTCGTCGTGTCG 4649
Qy 4321 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4380
Db 4650 ACCGATGCTCTCATGACTGGCTTTACCGGCGACTTCGACTCTGTGATAGACTGCAACACG 4709
Qy 4381 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4440
Db 4710 TGTGTCACTCAGACAGTCGATTTCAGCCTTGACCCTACCTTTACCATTGAGACAACCACG 4769
Qy 4441 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4500
Db 4770 CTCCCCCAGGATGCTGTCTCCAGGACTCAACGCCGGGGCAGGACTGGCAGGGGGAAGCCA 4829
Qy 4501 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4560
Db 4830 GGCATCTATAGATTTGTGGCACCGGGGGAGCGCCCCTCCGGCATGTTCGACTCGTCCGTC 4889
Qy 4561 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4620
Db 4890 CTCTGTGAGTGCTATGACGCGGGCTGTGCTTGGTATGAGCTCACGCCCGCCGAGACTACA 4949
Qy 4621 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 4680
Db 4950 GTTAGGCTACGAGCGTACATGAACACCCCGGGGCTTCCCGTGTGCCAGGACCATCTTGAA 5009
Qy 4681 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 4740
Db 5010 TTTTGGGAGGGCGTCTTTACGGGCCTCACTCATATAGATGCCCACTTTTTATCCCAGACA 5069
Qy 4741 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 4800
Db 5070 AAGCAGAGTGGGGAGAACTTTCCTTACCTGGTAGCGTACCAAGCCACCGTGTGCGCTAGG 5129
Qy 4801 GCCAAAGCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 4860
Db 5130 GCTCAAGCCCCTCCCCCATCGTGGGACCAGATGTGGAAGTGTTTGATCCGCCTTAAACCC 5189
Qy 4861 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 4920
Db 5190 ACCCTCCATGGGCCAACACCCCTGCTATACAGACTGGGCGCTGTTCAGAATGAAGTCACC 5249
Qy 4921 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 4980
Db 5250 CTGACGCACCCAATCACCAAATACATCATGACATGCATGTCGGCCGACCTGGAGGTCGTC 5309
Qy 4981 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5040
Db 5310 ACGAGCACCTGGGTGCTCGTTGGCGGCGTCCTGGCTGCTCTGGCCGCGTATTGCCTGTCA 5369
Qy 5041 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5100
Db 5370 ACAGGCTGCGTGGTCATAGTGGGCAGGATCGTCTTGTCCGGGAAGCCGGCAATTATACCT 5429
Qy 5101 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5160
Db 5430 GACAGGGAGGTTCTCTACCAGGAGTTCGATGAGATGGAAGAGTGCTCTCAGCACTTACCG 5489
Qy 5161 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5220
Db 5490 TACATCGAGCAAGGGATGATGCTCGCTGAGCAGTTCAAGCAGAAGGCCCTCGGCCTCCTG 5549
Qy 5221 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5280
Db 5550 CAGACCGCGTCCCGCCATGCAGAGGTTATCACCCCTGCTGTCCAGACCAACTGGCAGAAA 5609
Qy 5281 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5340
Db 5610 CTCGAGGTCTTTTGGGCGAAGCACATGTGGAATTTCATCAGTGGGATACAATACTTGGCG 5669
Qy 5341 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5400
Db 5670 GGCCTGTCAACGCTGCCTGGTAACCCCGCCATTGCTTCATTGATGGCTTTTACAGCTGCC 5729
Qy 5401 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5460
Db 5730 GTCACCAGCCCACTAACCACTGGCCAAACCCTCCTCTTCAACATATTGGGGGGGTGGGTG 5789
Qy 5461 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5520
Db 5790 GCTGCCCAGCTCGCCGCCCCCGGTGCCGCTACTGCCTTTGTGGGTGCTGGCCTAGCTGGC 5849
Qy 5521 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5580
Db 5850 GCCGCCATCGGCAGCGTTGGACTGGGGAAGGTCCTCGTGGACATTCTTGCAGGGTATGGC 5909
Qy 5581 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5640
Db 5910 GCGGGCGTGGCGGGAGCTCTTGTAGCATTCAAGATCATGAGCGGTGAGGTCCCCTCCACG 5969
Qy 5641 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 5700
Db 5970 GAGGACCTGGTCAATCTGCTGCCCGCCATCCTCTCGCCTGGAGCCCTTGTAGTCGGTGTG 6029
Qy 5701 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 5760
Db 6030 GTCTGCGCAGCAATACTGCGCCGGCACGTTGGCCCGGGCGAGGGGGCAGTGCAATGGATG 6089
Qy 5761 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 5820
Db 6090 AACCGGCTAATAGCCTTCGCCTCCCGGGGGAACCATGTTTCCCCCACGCACTACGTGCCG 6149

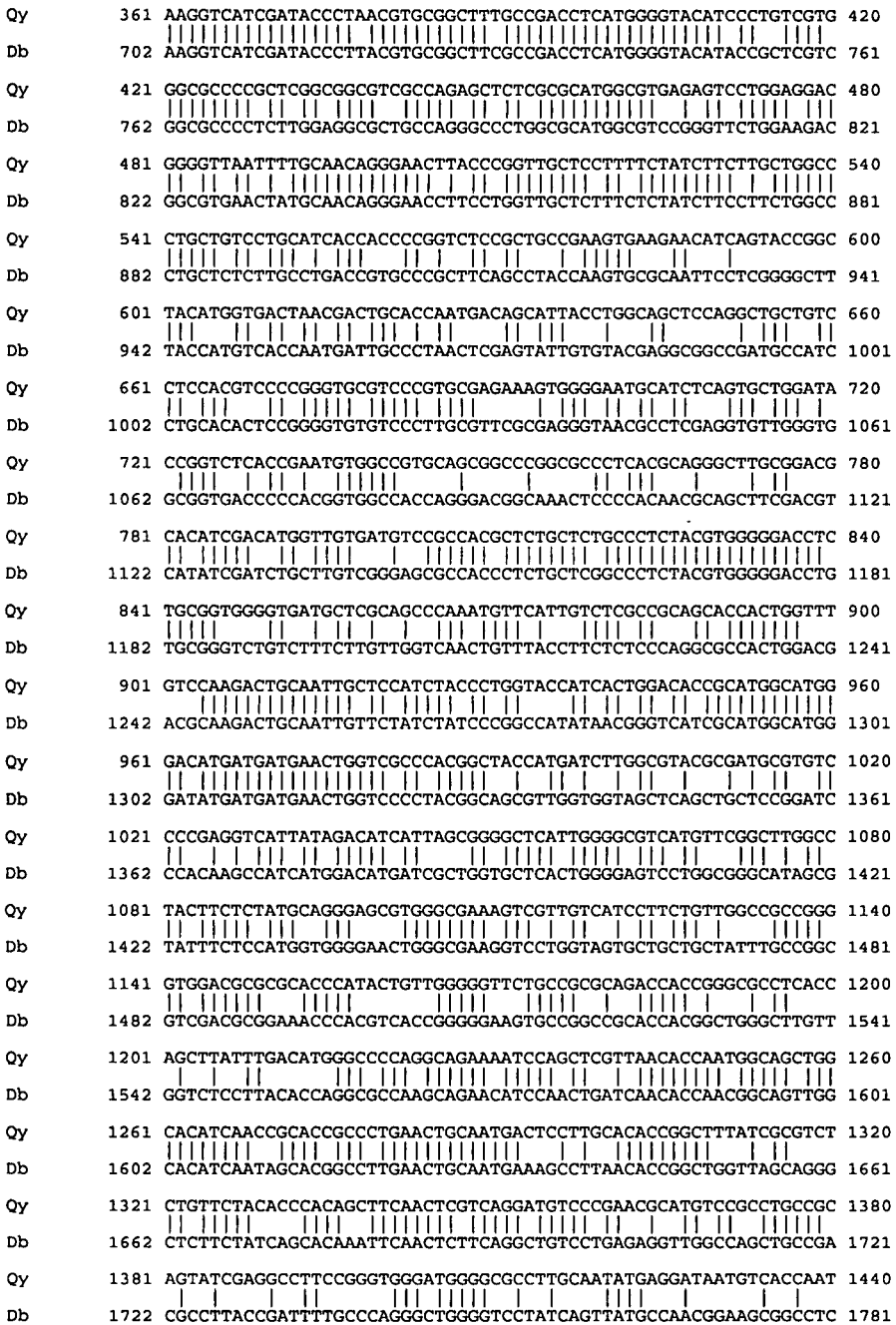


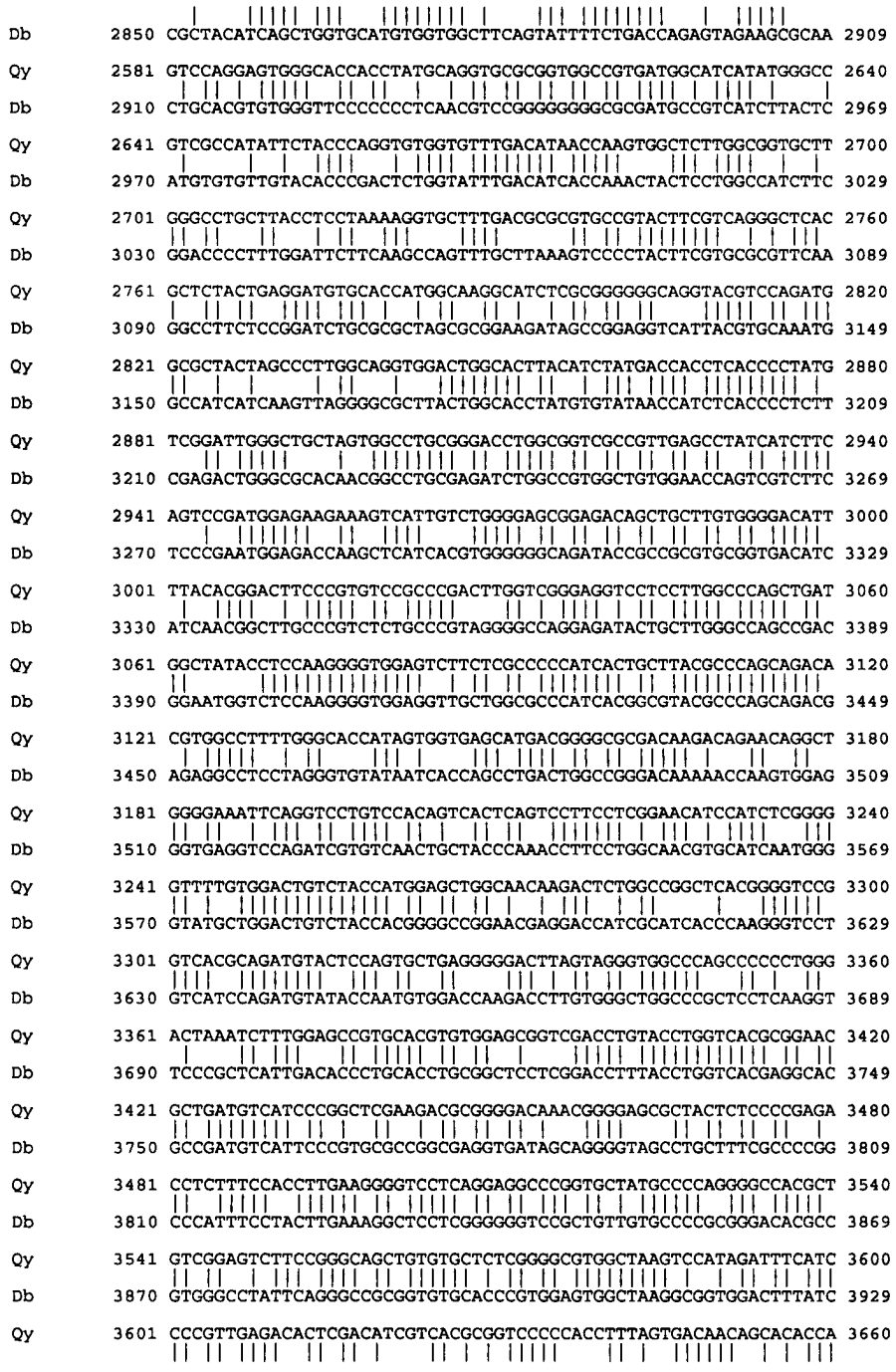


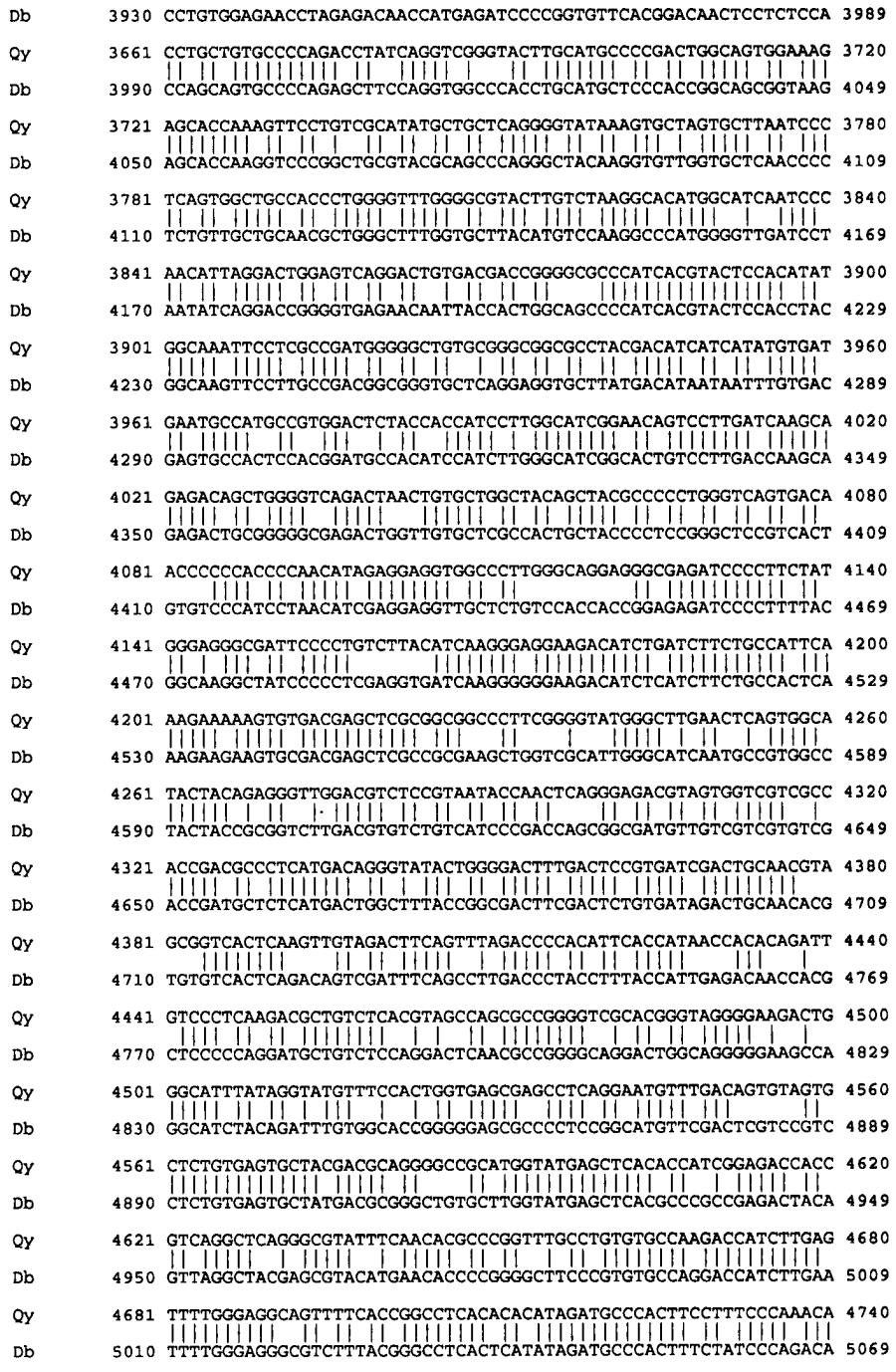
```
RESULT 7
US-08-811-566-5
; Sequence 5, Application US/08811566
; Patent No. 6127116
;  GENERAL INFORMATION:
;    APPLICANT:  Rice, Charles  et al.
;    TITLE OF INVENTION:  FUNCTIONAL DNA CLONE FOR HEPATITIS C
;    TITLE OF INVENTION:  VIRUS (HCV) AND USES THEREOF
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  David A. Jackson, Esq.
;      STREET:  411 Hackensack Ave, Continental Plaza, 4th
;      STREET:  Floor
;      CITY:  Hackensack
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/811,566
;      FILING DATE:  03-MAR-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jackson Esq., David A.
;      REGISTRATION NUMBER:  26,742
;      REFERENCE/DOCKET NUMBER:  1113-1-006
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-487-5800
;      TELEFAX:  201-343-1684
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12980 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-811-566-5

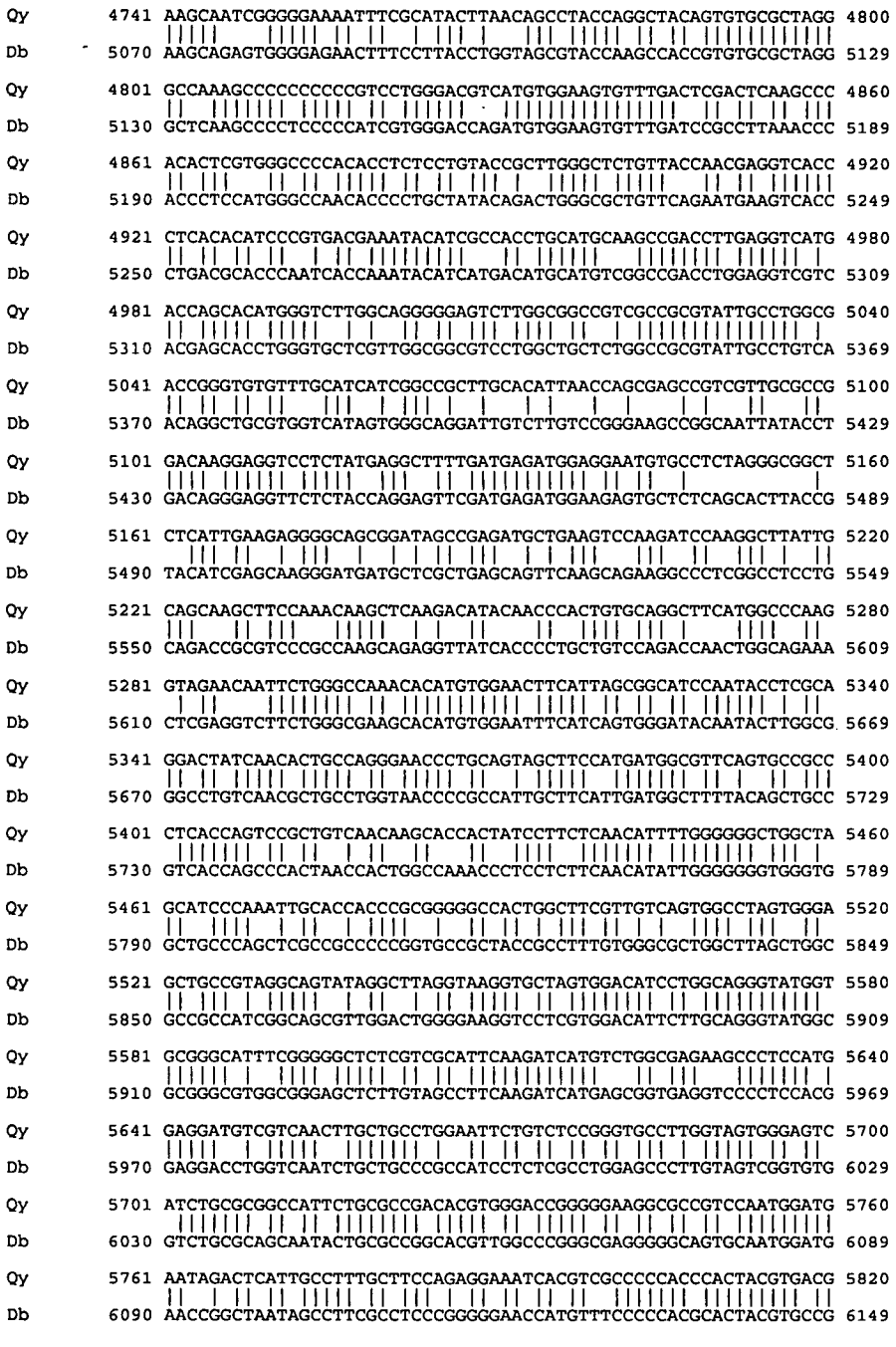
  Query Match             45.4%;  Score 4129.6;  DB 3;  Length 12980;
  Best Local Similarity   66.4%;  Pred. No. 0;
  Matches 6051;  Conservative    0;  Mismatches 2969;  Indels   90;  Gaps    5;
```

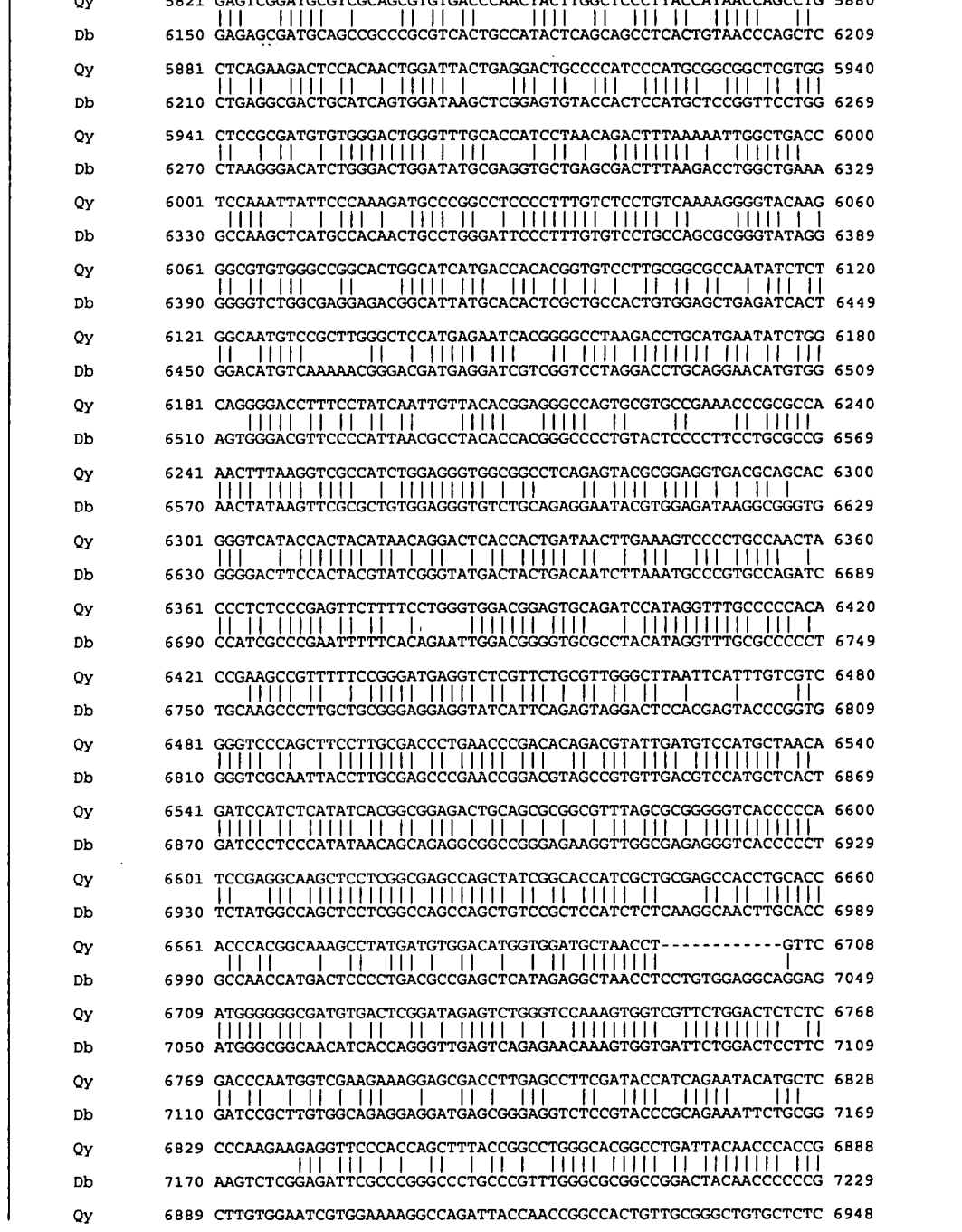


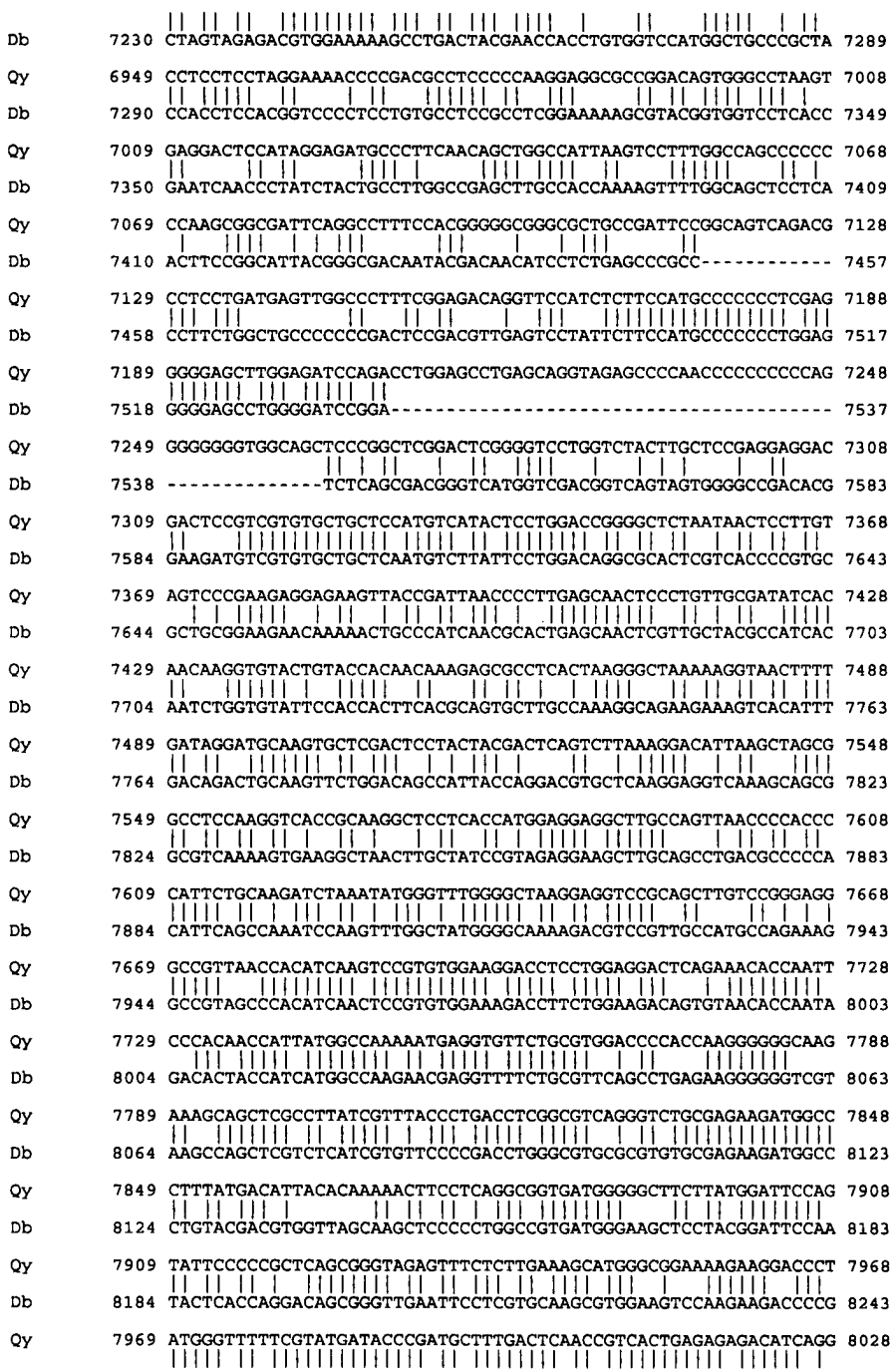

Qy 361 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 420
Db 702 AAGGTCATCGATACCCTTACGTGCGGCTTCGCCGACCTCATGGGGTACATACCGCTCGTC 761
Qy 421 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 480
Db 762 GGCGCCCCTCTTGGAGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAAGAC 821
Qy 481 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 540
Db 822 GGCGTGAACTATGCAACAGGGAACCTTCCTGGTTGCTCTTTCTCTATCTTCCTTCTGGCC 881
Qy 541 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 600
Db 882 CTGCTCTCTTGCCTGACCGTGCCCGCTTCAGCCTACCAAGTGCGCAATTCCTCGGGGCTT 941
Qy 601 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 660
Db 942 TACCATGTCACCAATGATTGCCCTAACTCGAGTATTGTGTACGAGGCGGCCGATGCCATC 1001
Qy 661 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 720
Db 1002 CTGCACACTCCGGGGGTGTGTCCCTTGCGTTCGCGAGGGTAACGCCTCGAGGTGTTGGGTG 1061
Qy 721 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 780
Db 1062 GCGGTGACCCCCACGGTGGCCACCAGGGACGGCAAACTCCCCACAACGCAGCTTCGACGT 1121
Qy 781 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 840
Db 1122 CATATCGATCTGCTTGTCGGGAGCGCCACCCTCTGCTCGGCCCTCTACGTGGGGGACCTG 1181
Qy 841 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 900
Db 1182 TGCGGGTCTGTCTTTCTTGTTGGTCAACTGTTTACCTTCTCTCCCAGGCGCCACTGGACG 1241
Qy 901 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 960
Db 1242 ACGCAAGACTGCAATTGTTCTATCTATCCCGGCCATATAACGGGTCATCGCATGGCATGG 1301
Qy 961 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1020
Db 1302 GATATGATGATGAACTGGTCCCCTACGGCAGCGTTGGTGGTAGCTCAGCTGCTCCGGATC 1361
Qy 1021 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1080
Db 1362 CCACAAGCCATCATGGACATGATCGCTGGTGCTCACTGGGGAGTCCTGGCGGGCATAGCG 1421
Qy 1081 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1140
Db 1422 TATTTCTCCATGGTGGGGAACTGGGCGAAGGTCCTGGTAGTGCTGCTGCTATTTGCCGGC 1481
Qy 1141 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1200
Db 1482 GTCGACGCGGAAACCCACGTCACCGGGGGAAGTGCCGGCCGCACCACGGCTGGGCTTGTT 1541
Qy 1201 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1260
Db 1542 GGTCTCCTTACACCAGGCGCCAAGCAGAACATCCAACTGATCAACACCAACGGCAGTTGG 1601
Qy 1261 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1320
Db 1602 CACATCAATAGCACGGCCTTGAACTGCAATGAAAGCCTTAACACCGGCTGGTTAGCAGGG 1661
Qy 1321 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1380
Db 1662 CTCTTCTATCAGCACAAATTCAACTCTTCAGGCTGTCCTGAGAGGTTGGCCAGCTGCCGA 1721
Qy 1381 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1440
Db 1722 CGCCTTACCGATTTTGCCCAGGGCTGGGGTCCTATCAGTTATGCCAACGGAAGCGGCCTC 1781
Qy 1441 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1500
Db 1782 GACGA------ACGCCCCTACTGCTGGCACTACCCTCCAAGACCTTGTGGCATTGTGCCC 1835
Qy 1501 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1560
Db 1836 GCAAAGAGCGTGTGTGGCCCGGTATATTGCTTCACTCCCAGCCCCGTGGTGGTGGGAACG 1895
Qy 1561 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1620
Db 1896 ACCGACAGGTCGGGCGCGCCTACCTACAGCTGGGGTGCAAATGATACGGATGTCTTCGTC 1955
Qy 1621 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 1680
Db 1956 CTTAACAACACCAGGCCACCGCTGGGCAATTGGTTCGGTTGTACCTGGATGAACTCAACT 2015
Qy 1681 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 1740
Db 2016 GGATTCACCAAAGTGTGCGGAGCGCCCCCTTGTGTCATCGGAGGGGTGGGCAACAACA-- 2073
Qy 1741 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 1800
Db 2074 ----CCTTGCTCTGCCCCACTGATTGTTTCCGCAAGCATCCGGAAGCCACATACTCTCGG 2129
Qy 1801 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 1860
Db 2130 TGCGGCTCCGGTCCCTGGATTACACCCAGGTGCATGGTCGACTACCCGTATAGGCTTTGG 2189
Qy 1861 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 1920
Db 2190 CACTATCCTTGTACCATCAATTACACCATATTCAAAGTCAGGATGTACGTGGGAGGGGTC 2249
Qy 1921 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 1980
Db 2250 GAGCACAGGCTGGAAGCGGCCTGCAACTGGACGCGGGGCGAACGCTGTGATCTGGAAGAC 2309
Qy 1981 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2040
Db 2310 AGGGACAGGTCCGAGCTCAGCCCATTGCTGCTGTCCACCACACAGTGGCAGGTCCTTCCG 2369
Qy 2041 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2100
Db 2370 TGTTCTTTCACGACCCTGCCAGCCTTGTCCACCGGCCTCATCCACCTCCACCAGAACATT 2429
Qy 2101 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2160
Db 2430 GTGGACGTGCAGTACTTGTACGGGGTAGGGTCAAGCATCGCGTCCTGGGCCATTAAGTGG 2489
Qy 2161 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2220
Db 2490 GAGTACGTCGTTCTCCTGTTCCTCCTGCTTGCAGACGCGCGCGTCTGCTCCTGCTTGTGG 2549
Qy 2221 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2280
Db 2550 ATGATGTTACTCATATCCCAAGCGGAGGCGGCTTTGGAGAACCTCGTAATACTCAATGCA 2609
Qy 2281 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATATTTTGTCATCTTTTTCGTGGCTGCTTGG 2340
Db 2610 GCATCCCTGGCCGGGACGCACGGTCTTGTGTCCTTCCTCGTGTTCTTCTGCTTTGCGTGG 2669
Qy 2341 TACATCAAGGGTCGGGTAGTCCCCTTAGCTACCTATTCCCTCACTGGCCTGTGGTCCTTT 2400
Db 2670 TATCTGAAGGGTAGGTGGGTGCCCGGAGCGGTCTACGCCTTCTACGGGATGTGGCCTCTC 2729
Qy 2401 AGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTGCATGGC 2460
Db 2730 CTCCTGCTCCTGCTGGCGTTGCCTCAGCGGGCATACGCACTGGACACGGAGGTGGCCGCG 2789
Qy 2461 CAGATAGGAGCGGCTCTGCTGGTAATGATCACTCTCTTTACTCTCACCCCCGGGTATAAG 2520
Db 2790 TCGTGTGGCGGCGTTGTTCTTGTCGGGTTAATGGCGCTGACTCTGTCGCCATATTACAAG 2849
Qy 2521 ACCCTTCTCAGCCGGTTTTTGTGGTGGTTGTGCTATCTTCTGACCCTGGGGGAAGCTATG 2580


Db 2850 CGCTACATCAGCTGGTGCATGTGGTGGCTTCAGTATTTTCTGACCAGAGTAGAAGCGCAA 2909
Qy 2581 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGCATCATATGGGCC 2640
Db 2910 CTGCACGTGTGGGTTCCCCCCCTCAACGTCCGGGGGGGGCGCGATGCCGTCATCTTACTC 2969
Qy 2641 GTCGCCATATTCTACCCAGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 2700
Db 2970 ATGTGTGTTGTACACCCGACTCTGGTATTTGACATCACCAAACTACTCCTGGCCATCTTC 3029
Qy 2701 GGGCCTGCTTACCTCCTAAAAGGTGCTTTGACGCGCGTGCCGTACTTCGTCAGGGCTCAC 2760
Db 3030 GGACCCCTTTGGATTCTTCAAGCCAGTTTGCTTAAAGTCCCCTACTTCGTGCGCGTTCAA 3089
Qy 2761 GCTCTACTGAGGATGTGCACCATGGCAAGGCATCTCGCGGGGGGCAGGTACGTCCAGATG 2820
Db 3090 GGCCTTCTCCGGATCTGCGCGCTAGCGCGGAAGATAGCCGGAGGTCATTACGTGCAAATG 3149
Qy 2821 GCGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 2880
Db 3150 GCCATCATCAAGTTAGGGGCGCTTACTGGCACCTATGTGTATAACCATCTCACCCCTCTT 3209
Qy 2881 TCGGATTGGGCTGCTAGTGGCCTGCGGGACCTGGCGGTCGCCGTTGAGCCTATCATCTTC 2940
Db 3210 CGAGACTGGGCGCACAACGGCCTGCGAGATCTGGCCGTGGCTGTGGAACCAGTCGTCTTC 3269
Qy 2941 AGTCCGATGGAGAAGAAAGTCATTGTCTGGGGAGCGGAGACAGCTGCTTGTGGGGACATT 3000
Db 3270 TCCCGAATGGAGACCAAGCTCATCACGTGGGGGGCAGATACCGCCGCGTGCGGTGACATC 3329
Qy 3001 TTACACGGACTTCCCGTGTCCGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3060
Db 3330 ATCAACGGCTTGCCCGTCTCTGCCCGTAGGGGCCAGGAGATACTGCTTGGGCCAGCCGAC 3389
Qy 3061 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3120
Db 3390 GGAATGGTCTCCAAGGGGTGGAGGTTGCTGGCGCCCATCACGGCGTACGCCCAGCAGACG 3449
Qy 3121 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCT 3180
Db 3450 AGAGGCCTCCTAGGGTGTATAATCACCAGCCTGACTGGCCGGGACAAAAACCAAGTGGAG 3509
Qy 3181 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3240
Db 3510 GGTGAGGTCCAGATCGTGTCAACTGCTACCCAAACCTTCCTGGCAACGTGCATCAATGGG 3569
Qy 3241 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3300
Db 3570 GTATGCTGGACTGTCTACCACGGGGCCGGAACGAGGACCATCGCATCACCCAAGGGTCCT 3629
Qy 3301 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGG 3360
Db 3630 GTCATCCAGATGTATACCAATGTGGACCAAGACCTTGTGGGCTGGCCCGCTCCTCAAGGT 3689
Qy 3361 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3420
Db 3690 TCCCGCTCATTGACACCCTGCACCTGCGGCTCCTCGGACCTTTACCTGGTCACGAGGCAC 3749
Qy 3421 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3480
Db 3750 GCCGATGTCATTCCCGTGCGCCGGCGAGGTGATAGCAGGGGTAGCCTGCTTTCGCCCCGG 3809
Qy 3481 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3540
Db 3810 CCCATTTCCTACTTGAAAGGCTCCTCGGGGGGTCCGCTGTTGTGCCCCGCGGGACACGCC 3869
Qy 3541 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3600
Db 3870 GTGGGCCTATTCAGGGCCGCGGTGTGCACCCGTGGAGTGGCTAAGGCGGTGGACTTTATC 3929
Qy 3601 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 3660


Db 3930 CCTGTGGAGAACCTAGAGACAACCATGAGATCCCCGGTGTTCACGGACAACTCCTCTCCA 3989
Qy 3661 CCTGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 3720
Db 3990 CCAGCAGTGCCCCAGAGCTTCCAGGTGGCCCACCTGCATGCTCCCACCGGCAGCGGTAAG 4049
Qy 3721 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 3780
Db 4050 AGCACCAAGGTCCCGGCTGCGTACGCAGCCCAGGGCTACAAGGTGTTGGTGCTCAACCCC 4109
Qy 3781 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 3840
Db 4110 TCTGTTGCTGCAACGCTGGGCTTTGGTGCTTACATGTCCAAGGCCCATGGGGTTGATCCT 4169
Qy 3841 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 3900
Db 4170 AATATCAGGACCGGGGTGAGAACAATTACCACTGGCAGCCCCATCACGTACTCCACCTAC 4229
Qy 3901 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 3960
Db 4230 GGCAAGTTCCTTGCCGACGGCGGGTGCTCAGGAGGTGCTTATGACATAATAATTTGTGAC 4289
Qy 3961 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4020
Db 4290 GAGTGCCACTCCACGGATGCCACATCCATCTTGGGCATCGGCACTGTCCTTGACCAAGCA 4349
Qy 4021 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4080
Db 4350 GAGACTGCGGGGGCGAGACTGGTTGTGCTCGCCACTGCTACCCCTCCGGGCTCCGTCACT 4409
Qy 4081 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4140
Db 4410 GTGTCCCATCCTAACATCGAGGAGGTTGCTCTGTCCACCACCGGAGAGATCCCCTTTTAC 4469
Qy 4141 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4200
Db 4470 GGCAAGGCTATCCCCCTCGAGGTGATCAAGGGGGGAAGACATCTCATCTTCTGCCACTCA 4529
Qy 4201 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4260
Db 4530 AAGAAGAAGTGCGACGAGCTCGCCGCGAAGCTGGTCGCATTGGGCATCAATGCCGTGGCC 4589
Qy 4261 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4320
Db 4590 TACTACCGCGGTCTTGACGTGTCTGTCATCCCGACCAGCGGCGATGTTGTCGTCGTGTCG 4649
Qy 4321 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4380
Db 4650 ACCGATGCTCTCATGACTGGCTTTACCGGCGACTTCGACTCTGTGATAGACTGCAACACG 4709
Qy 4381 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4440
Db 4710 TGTGTCACTCAGACAGTCGATTTCAGCCTTGACCCTACCTTTACCATTGAGACAACCACG 4769
Qy 4441 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4500
Db 4770 CTCCCCCAGGATGCTGTCTCCAGGACTCAACGCCGGGGCAGGACTGGCAGGGGGAAGCCA 4829
Qy 4501 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4560
Db 4830 GGCATCTACAGATTTGTGGCACCGGGGGAGCGCCCCTCCGGCATGTTCGACTCGTCCGTC 4889
Qy 4561 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4620
Db 4890 CTCTGTGAGTGCTATGACGCGGGCTGTGCTTGGTATGAGCTCACGCCCGCCGAGACTACA 4949
Qy 4621 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 4680
Db 4950 GTTAGGCTACGAGCGTACATGAACACCCCGGGGCTTCCCGTGTGCCAGGACCATCTTGAA 5009
Qy 4681 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 4740
Db 5010 TTTTGGGAGGGCGTCTTTACGGGCCTCACTCATATAGATGCCCACTTTCTATCCCAGACA 5069






Db 7230 CTAGTAGAGACGTGGAAAAAGCCTGACTACGAACCACCTGTGGTCCATGGCTGCCCGCTA 7289
Qy 6949 CCTCCTCCTAGGAAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGT 7008
Db 7290 CCACCTCCACGGTCCCCTCCTGTGCCTCCGCCTCGGAAAAAGCGTACGGTGGTCCTCACC 7349
Qy 7009 GAGGACTCCATAGGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCC 7068
Db 7350 GAATCAACCCTATCTACTGCCTTGGCCGAGCTTGCCACCAAAAGTTTTGGCAGCTCCTCA 7409
Qy 7069 CCAAGCGGCGATTCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACG 7128
Db 7410 ACTTCCGGCATTACGGGCGACAATACGACAACATCCTCTGAGCCCGCC------------ 7457
Qy 7129 CCTCCTGATGAGTTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCTCGAG 7188
Db 7458 CCTTCTGGCTGCCCCCCCGACTCCGACGTTGAGTCCTATTCTTCCATGCCCCCCCTGGAG 7517
Qy 7189 GGGGAGCTTGGAGATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCAG 7248
Db 7518 GGGGAGCCTGGGGATCCGGA-- 7537
Qy 7249 GGGGGGGTGGCAGCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGAC 7308
Db 7538 -------------TCTCAGCGACGGGTCATGGTCGACGGTCAGTAGTGGGGCCGACACG 7583
Qy 7309 GACTCCGTCGTGTGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGT 7368
Db 7584 GAAGATGTCGTGTGCTGCTCAATGTCTTATTCCTGGACAGGCGCACTCGTCACCCCGTGC 7643
Qy 7369 AGTCCCGAAGAGGAGAAGTTACCGATTAACCCCTTGAGCAACTCCCTGTTGCGATATCAC 7428
Db 7644 GCTGCGGAAGAACAAAAACTGCCCATCAACGCACTGAGCAACTCGTTGCTACGCCATCAC 7703
Qy 7429 AACAAGGTGTACTGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTT 7488
Db 7704 AATCTGGTGTATTCCACCACTTCACGCAGTGCTTGCCAAAGGCAGAAGAAAGTCACATTT 7763
Qy 7489 GATAGGATGCAAGTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCG 7548
Db 7764 GACAGACTGCAAGTTCTGGACAGCCATTACCAGGACGTGCTCAAGGAGGTCAAAGCAGCG 7823
Qy 7549 GCCTCCAAGGTCACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCC 7608
Db 7824 GCGTCAAAAGTGAAGGCTAACTTGCTATCCGTAGAGGAAGCTTGCAGCCTGACGCCCCCA 7883
Qy 7609 CATTCTGCAAGATCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGG 7668
Db 7884 CATTCAGCCAAATCCAAGTTTGGCTATGGGGCAAAAGACGTCCGTTGCCATGCCAGAAAG 7943
Qy 7669 GCCGTTAACCACATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATT 7728
Db 7944 GCCGTAGCCCACATCAACTCCGTGTGGAAAGACCTTCTGGAAGACAGTGTAACACCAATA 8003
Qy 7729 CCCACAACCATTATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAG 7788
Db 8004 GACACTACCATCATGGCCAAGAACGAGGTTTTCTGCGTTCAGCCTGAGAAGGGGGGTCGT 8063
Qy 7789 AAAGCAGCTCGCCTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCC 7848
Db 8064 AAGCCAGCTCGTCTCATCGTGTTCCCCGACCTGGGCGTGCGCGTGTGCGAGAAGATGGCC 8123
Qy 7849 CTTTATGACATTACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAG 7908
Db 8124 CTGTACGACGTGGTTAGCAAGCTCCCCCTGGCCGTGATGGGAAGCTCCTACGGATTCCAA 8183
Qy 7909 TATTCCCCCGCTCAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCT 7968
Db 8184 TACTCACCAGGACAGCGGGTTGAATTCCTCGTGCAAGCGTGGAAGTCCAAGAAGACCCCG 8243
Qy 7969 ATGGGTTTTTCGTATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGG 8028
Db 8244 ATGGGGTTCTCGTATGATACCCGCTGTTTTGACTCCACAGTCACTGAGAGCGACATCCGT 8303
Qy 8029 ACTGAGGAGTCCATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACAC 8088
Db 8304 ACGGAGGAGGCAATTTACCAATGTTGTGACCTGGACCCCCAAGCCCGCGTGGCCATCAAG 8363
Qy 8089 TCGCTAACTGAGAGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGC 8148
Db 8364 TCCCTCACTGAGAGGCTTTATGTTGGGGGCCCTCTTACCAATTCAAGGGGGGAAAACTGC 8423
Qy 8149 GGGTACAGGCGTTGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACA 8208
Db 8424 GGCTACCGCAGGTGCCGCGCGAGCGGCGTACTGACAACTAGCTGTGGTAACACCCTCACT 8483
Qy 8209 TGCTACGTGAAAGCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTG 8268
Db 8484 TGCTACATCAAGGCCCGGGCAGCCTGTCGAGCCGCAGGGCTCCAGGACTGCACCATGCTC 8543
Qy 8269 GTATGCGGCGATGACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGG 8328
Db 8544 GTGTGTGGCGACGACTTAGTCGTTATCTGTGAAAGTGCGGGGGTCCAGGAGGACGCGGCG 8603
Qy 8329 AACCTGAGAGCCTTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCC 8388
Db 8604 AGCCTGAGAGCCTTCACGGAGGCTATGACCAGGTACTCCGCCCCCCCCGGGGACCCCCCA 8663
Qy 8389 AGACCGGAGTATGATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTG 8448
Db 8664 CAACCAGAATACGACTTGGAGCTTATAACATCATGCTCCTCCAACGTGTCAGTCGCCCAC 8723
Qy 8449 GGCCCACAAGGCCGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGG 8508
Db 8724 GACGGCGCTGGAAAGAGGGTCTACTACCTTACCCGTGACCCTACAACCCCCCTCGCGAGA 8783
Qy 8509 GCTGCCTGGGAAACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAG 8568
Db 8784 GCCGCGTGGGAGACAGCAAGACACACTCCAGTCAATTCCTGGCTAGGCAACATAATCATG 8843
Qy 8569 TACGCCCCGACCATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATG 8628
Db 8844 TTTGCCCCCACACTGTGGGCGAGGATGATACTGATGACCCATTTCTTTAGCGTCCTCATA 8903
Qy 8629 GCTCAAGACACGCTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTG 8688
Db 8904 GCCAGGGATCAGCTTGAACAGGCTCTTAACTGTGAGATCTACGGAGCCTGCTACTCCATA 8963
Qy 8689 AGTCCCTTGGACCTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTG 8748
Db 8964 GAACCACTGGATCTACCTCCAATCATTCAAAGACTCCATGGCCTCAGCGCATTTTCACTC 9023
Qy 8749 CACACATACACTCCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCG 8808
Db 9024 CACAGTTACTCTCCAGGTGAAATCAATAGGGTGGCCGCATGCCTCAGAAAACTTGGGGTC 9083
Qy 8809 CCACCCCTCAGAGCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGT 8868
Db 9084 CCGCCCTTGCGAGCTTGGAGACACCGGGCCCGGAGCGTCCGCGCTAGGCTTCTGTCCAGA 9143
Qy 8869 GGGGGGAGAGCGGCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTC 8928
Db 9144 GGAGGCAGGGCTGCCATATGTGGCAAGTACCTCTTCAACTGGGCAGTAAGAACAAAGCTC 9203
Qy 8929 AAACTCACTCCATTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGC 8988
Db 9204 AAACTCACTCCAATAGCGGCCGCTGGCCGGCTGGACTTGTCCGGTTGGTTCACGGCTGGC 9263
Qy 8989 GCCGGCGGGGGCGACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTT 9048
Db 9264 TACAGCGGGGGAGACATTTATCACAGCGTGTCTCATGCCCGGCCCCGCTGGTTCTGGTTT 9323
Qy 9049 GGCCTACTCCTACTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCG 9098
Db 9324 TGCCTACTCCTGCTCGCTGCAGGGGTAGGCATCTACCTCCTCCCCAACCG 9373

```
RESULT 8
US-09-034-756-5
; Sequence 5, Application US/09034756
; Patent No. 6392028
;   GENERAL INFORMATION:
;        APPLICANT: RICE, CHARLES  et al.
;        TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
;                            VIRUS (HCV) AND USES THEREOF
;        NUMBER OF SEQUENCES: 21
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: HOWELL & HAFERKAMP, L.C.
;             STREET: 7733 FORSYTH BLVD., SUITE 1400
;             CITY: ST. LOUIS
;             STATE: MO
;             COUNTRY: USA
;             ZIP: 63105
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/034,756
;             FILING DATE: 04-May-1998
;             CLASSIFICATION: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: HOLLAND, DONALD R.
;             REGISTRATION NUMBER: 35,197
;             REFERENCE/DOCKET NUMBER: 6029-4831
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 314-727-5188
;             TELEFAX: 314-727-6092
;   INFORMATION FOR SEQ ID NO: 5:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 12980 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        MOLECULE TYPE: cDNA
;        HYPOTHETICAL: NO
;        ANTI-SENSE: NO
;        SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-034-756-5

  Query Match             45.4%;  Score 4129.6;  DB 3;  Length 12980;
  Best Local Similarity   66.4%;  Pred. No. 0;
  Matches 6051;  Conservative    0;  Mismatches 2969;  Indels   90;  Gaps    5;

Qy          1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 60
              |||||||| |||||||||||||||||||||||||| | ||||||||||||||||||||
Db        342 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGTCGCCCACAG 401

Qy         61 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 120
              ||||| ||||| ||||| ||||| |||||||||||| ||||| ||||||||||||||||||
Db        402 GACGTCAAGTTCCCGGGTGGCGGTCAGATCGTTGGTGGAGTTTACTTGTTGCCGCGCAGG 461

Qy        121 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 180
              ||||| || |||||||||||||||| |||||||||| |||||||| || || || ||
Db        462 GGCCCTAGATTGGGTGTGCGCGCGACGAGGAAGACTTCCGAGCGGTCGCAACCTCGAGGT 521

Qy        181 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 240
              || || ||||| ||||| || |  || ||  ||   ||||   ||||||    | || ||
Db        522 AGACGTCAGCCTATCCCCAAGGCACGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGG 581

Qy        241 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 300
              ||||| |||||||| || || ||||||||    ||| ||||| ||||||||||||||| |||
Db        582 TACCCTTGGCCCCTCTATGGCAATGAGGGTTGCGGGTGGGCGGGATGGCTCCTGTCTCCC 641

Qy        301 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 360
              || || || || ||    |||||||||| |||||||||| ||||||||||||   |||||
Db        642 CGTGGCTCTCGGCCTAGCTGGGGCCCCACAGACCCCCGGCGTAGGTCGCGCAATTTGGGT 701

Qy        361 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 420
              |||||||||||||||| |||||||||| |||||||||||||||||||| ||  ||||
Db        702 AAGGTCATCGATACCCTTACGTGCGGCTTCGCCGACCTCATGGGGTACATACCGCTCGTC 761

Qy        421 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 480
              ||||||| || || |||| ||||| || || |||||||||| | || ||||| |||
Db        762 GGCGCCCCTCTTGGAGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAAGAC 821

Qy        481 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 540
              || || || | |||||||||||||| | || |||||||| || |||||||| | ||||||
Db        822 GGCGTGAACTATGCAACAGGGAACCTTCCTGGTTGCTCTTTCTCTATCTTCCTTCTGGCC 881

Qy        541 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 600
              ||||| || ||| | |||   || |  || ||   | |||||   ||   |
Db        882 CTGCTCTCTTGCCTGACCGTGCCCGCTTCAGCCTACCAAGTGCGCAATTCCTCGGGGCTT 941

Qy        601 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 660
              |||   || || || || ||| | ||    || |||   |   ||      | |||  ||
Db        942 TACCATGTCACCAATGATTGCCCTAACTCGAGTATTGTGTACGAGGCGGCCGATGCCATC 1001

Qy        661 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 720
              || |||   || ||||| ||||| ||||      | ||| || || ||   ||| ||| |
Db       1002 CTGCACACTCCGGGGTGTGTCCCTTGCGTTCGCGAGGGTAACGCCTCGAGGTGTTGGGTG 1061

Qy        721 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 780
              ||||  | || |  ||||||     |   |    || |||   |   | ||
Db       1062 GCGGTGACCCCCACGGTGGCCACCAGGGACGGCAAACTCCCCACAACGCAGCTTCGACGT 1121

Qy        781 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 840
              || |||||  || ||||   | |||||| |||||||| ||||||||||||||||||||
Db       1122 CATATCGATCTGCTTGTCGGGAGCGCCACCCTCTGCTCGGCCCTCTACGTGGGGGACCTG 1181

Qy        841 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 900
              |||||    ||  | || |  | ||| |||| |   |||| ||   || |||||||
Db       1182 TGCGGGTCTGTCTTTCTTGTTGGTCAACTGTTTACCTTCTCTCCCAGGCGCCACTGGACG 1241

Qy        901 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 960
              |||||||||||||| || |||| || ||    || || || || ||||||||||||
Db       1242 ACGCAAGACTGCAATTGTTCTATCTATCCCGGCCATATAACGGGTCATCGCATGGCATGG 1301

Qy        961 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1020
              || |||||||||||||||| || |||||  |  || |  | ||  |  | | ||  ||
Db       1302 GATATGATGATGAACTGGTCCCCTACGGCAGCGTTGGTGGTAGCTCAGCTGCTCCGGATC 1361

Qy       1021 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1080
              ||  | | ||| || |||||  ||    || |||| |||| ||| ||   ||| | ||
Db       1362 CCACAAGCCATCATGGACATGATCGCTGGTGCTCACTGGGGAGTCCTGGCGGGCATAGCG 1421

Qy       1081 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1140
              || |||| ||| |||     |||||||| ||| | ||  | || ||| |    |||||
Db       1422 TATTTCTCCATGGTGGGGAACTGGGCGAAGGTCCTGGTAGTGCTGCTGCTATTTGCCGGC 1481

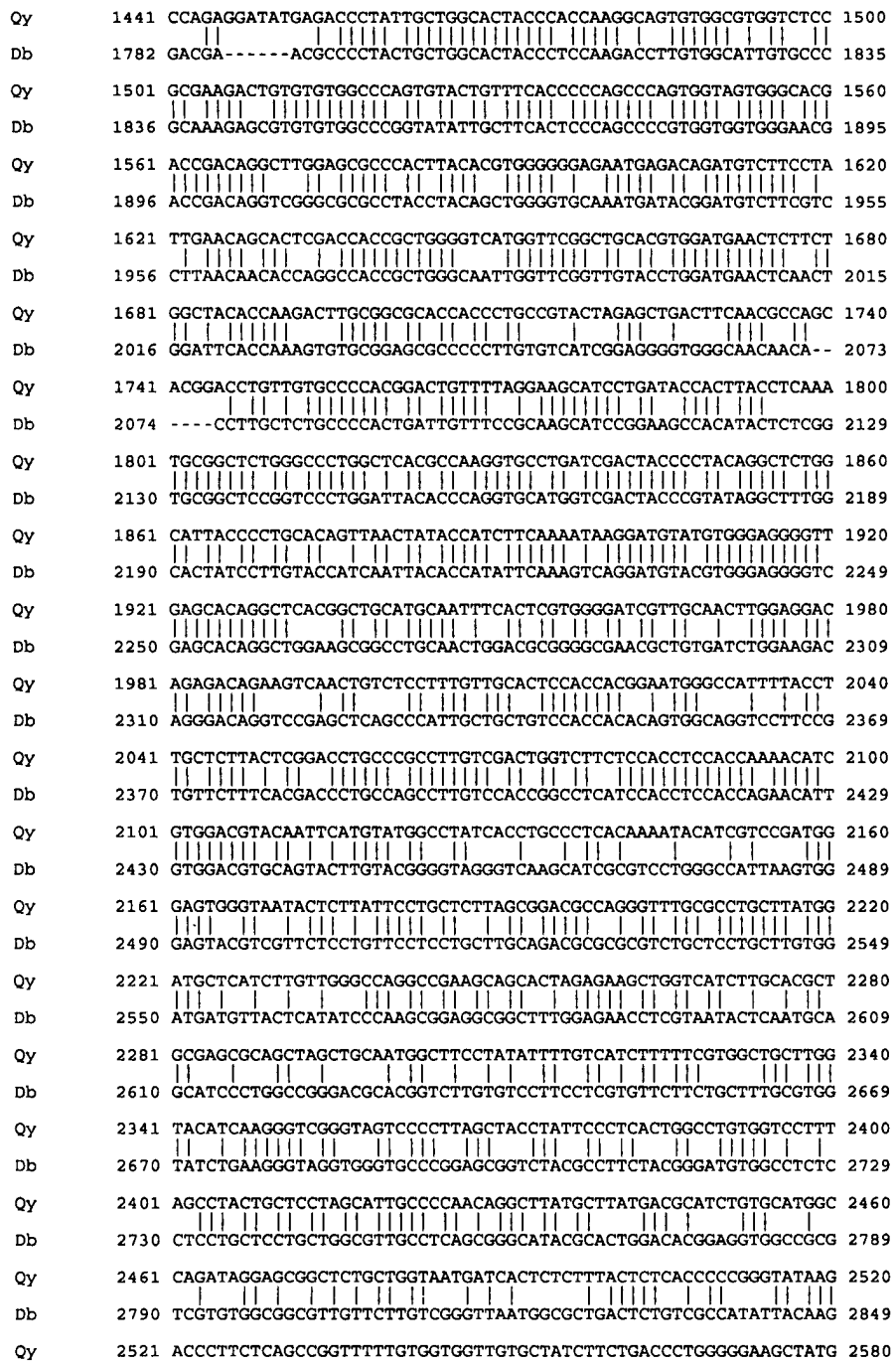
Qy       1141 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1200
              || ||||||  |||||        |||||  |||||  |  |||||  |   | ||
Db       1482 GTCGACGCGGAAACCCACGTCACCGGGGGAAGTGCCGGCCGCACCACGGCTGGGCTTGTT 1541

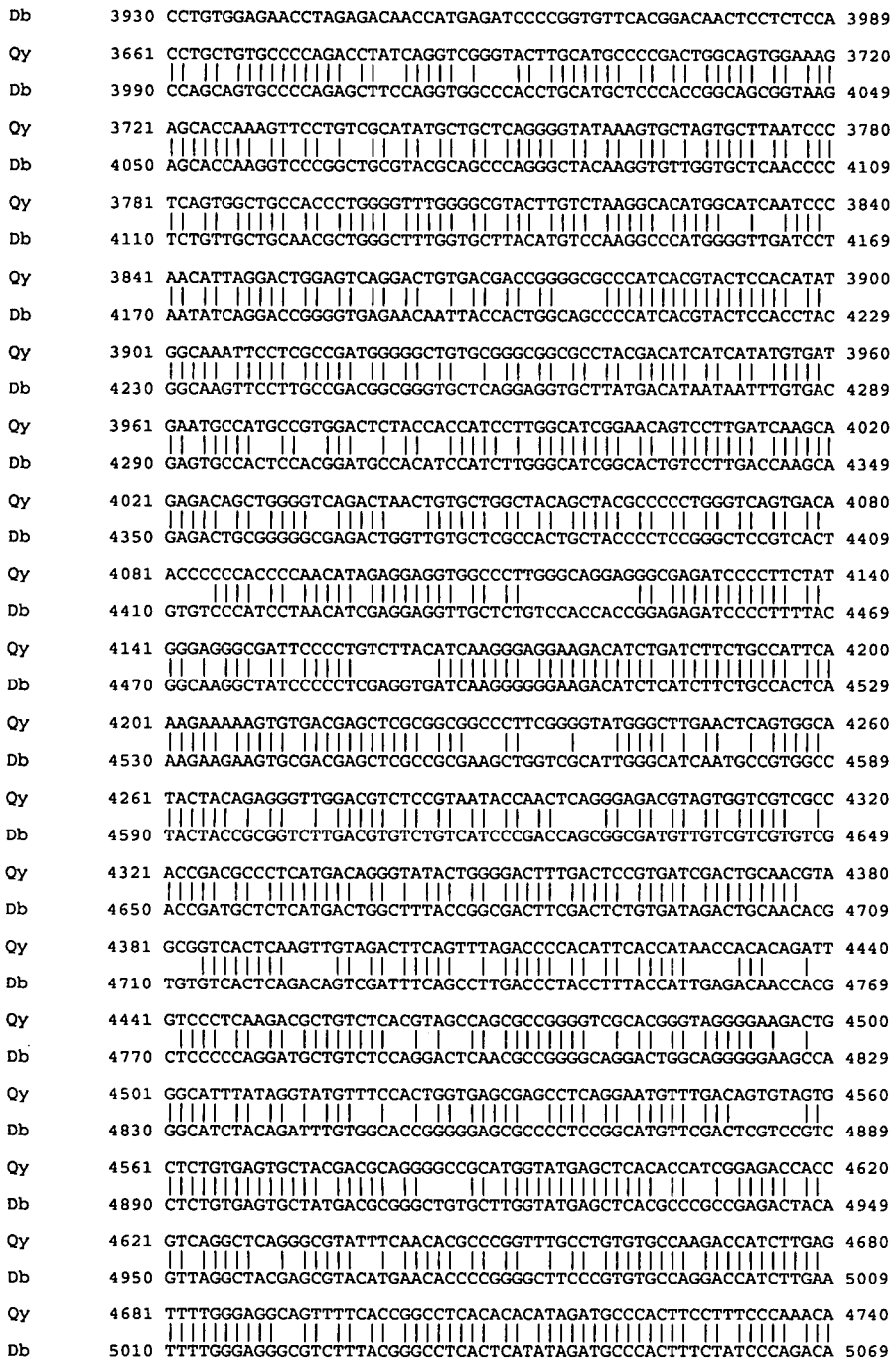
Qy       1201 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1260
              |  |  ||     ||| ||| |||||  |||| ||  | |||||||  ||||| |||
Db       1542 GGTCTCCTTACACCAGGCGCCAAGCAGAACATCCAACTGATCAACACCAACGGCAGTTGG 1601

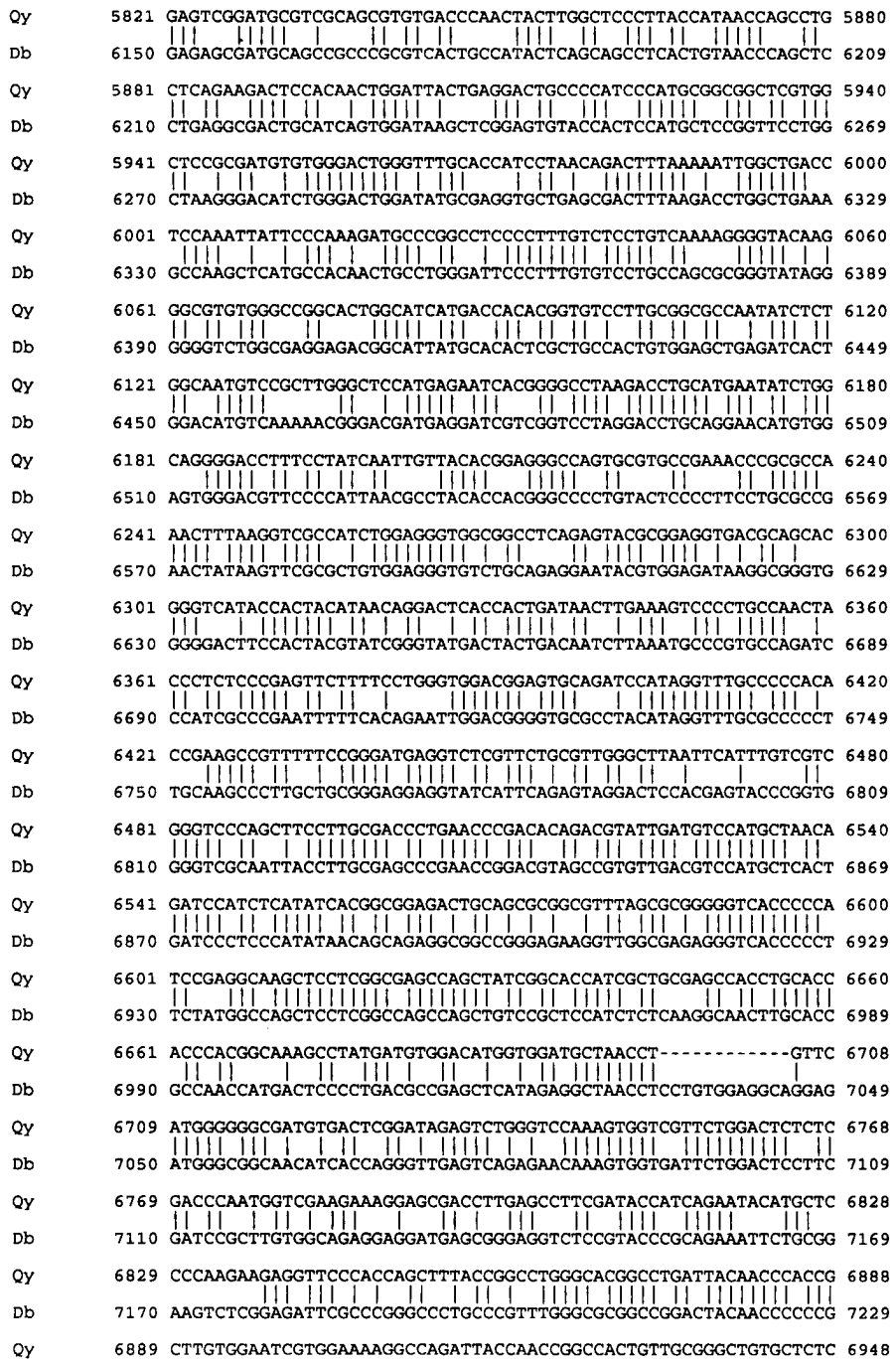
Qy       1261 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1320
              |||||||  |||| ||| ||||||||||||   |  |  |||||||||   | ||
Db       1602 CACATCAATAGCACGGCCTTGAACTGCAATGAAAGCCTTAACACCGGCTGGTTAGCAGGG 1661

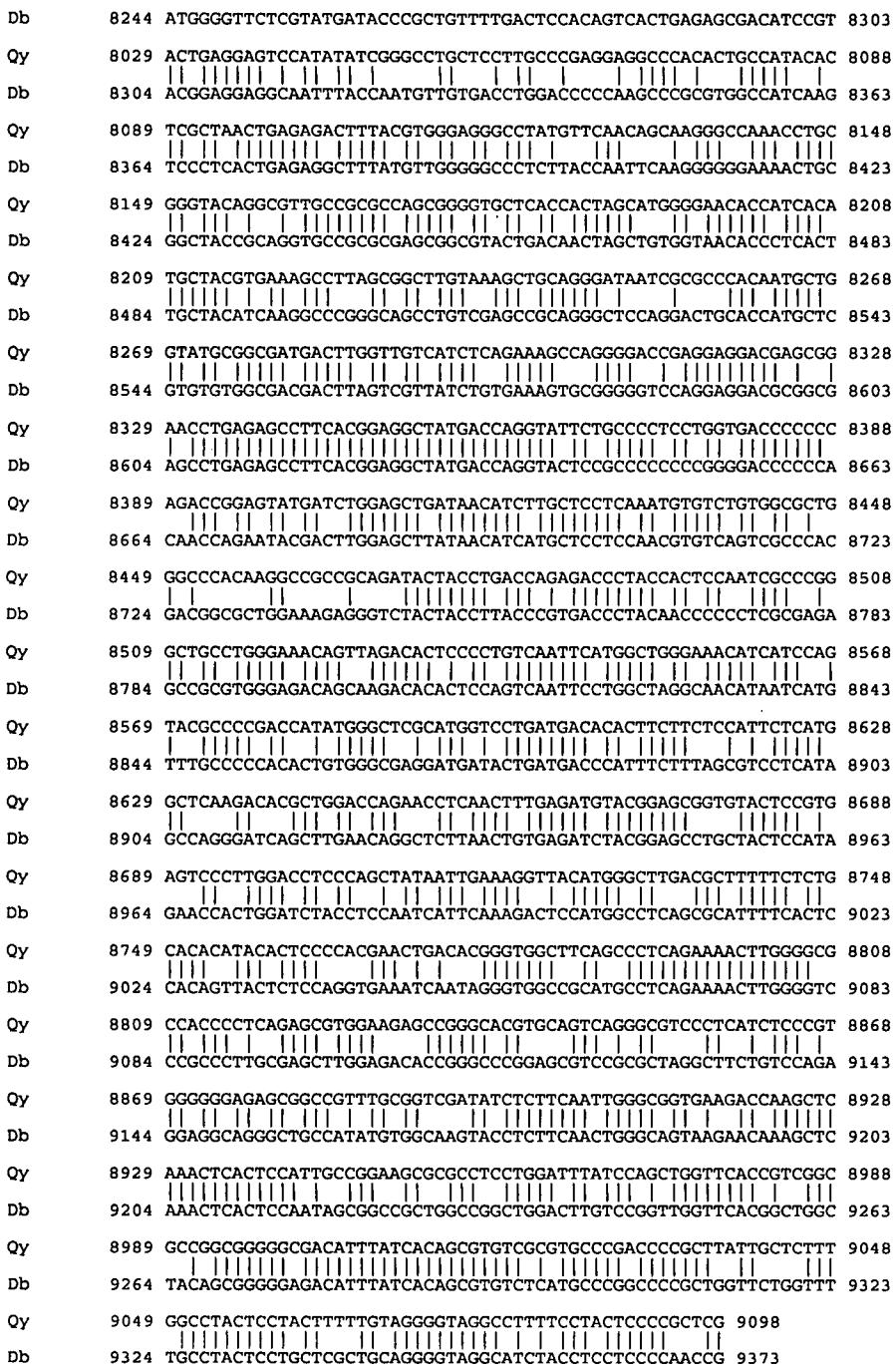
Qy       1321 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1380
              || |||||    ||||  |||||||| |||||  |||||  ||   |  || ||  ||||||
Db       1662 CTCTTCTATCAGCACAAATTCAACTCTTCAGGCTGTCCTGAGAGGTTGGCCAGCTGCCGA 1721

Qy       1381 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1440
              |  |     |  ||      ||| |||||  |  |    ||||   |      |  |
Db       1722 CGCCTTACCGATTTTGCCCAGGGCTGGGGTCCTATCAGTTATGCCAACGGAAGCGGCCTC 1781
```


Qy 1441 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1500
Db 1782 GACGA------ACGCCCCTACTGCTGGCACTACCCTCCAAGACCTTGTGGCATTGTGCCC 1835
Qy 1501 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1560
Db 1836 GCAAAGAGCGTGTGTGGCCCGGTATATTGCTTCACTCCCAGCCCCGTGGTGGTGGGAACG 1895
Qy 1561 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1620
Db 1896 ACCGACAGGTCGGGCGCGCCTACCTACAGCTGGGGTGCAAATGATACGGATGTCTTCGTC 1955
Qy 1621 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 1680
Db 1956 CTTAACAACACCAGGCCACCGCTGGGCAATTGGTTCGGTTGTACCTGGATGAACTCAACT 2015
Qy 1681 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 1740
Db 2016 GGATTCACCAAAGTGTGCGGAGCGCCCCCTTGTGTCATCGGAGGGGTGGGCAACAACA-- 2073
Qy 1741 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 1800
Db 2074 ----CCTTGCTCTGCCCCACTGATTGTTTCCGCAAGCATCCGGAAGCCACATACTCTCGG 2129
Qy 1801 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 1860
Db 2130 TGCGGCTCCGGTCCCTGGATTACACCCAGGTGCATGGTCGACTACCCGTATAGGCTTTGG 2189
Qy 1861 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 1920
Db 2190 CACTATCCTTGTACCATCAATTACACCATATTCAAAGTCAGGATGTACGTGGGAGGGGTC 2249
Qy 1921 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 1980
Db 2250 GAGCACAGGCTGGAAGCGGCCTGCAACTGGACGCGGGGCGAACGCTGTGATCTGGAAGAC 2309
Qy 1981 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2040
Db 2310 AGGGACAGGTCCGAGCTCAGCCCATTGCTGCTGTCCACCACACAGTGGCAGGTCCTTCCG 2369
Qy 2041 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2100
Db 2370 TGTTCTTTCACGACCCTGCCAGCCTTGTCCACCGGCCTCATCCACCTCCACCAGAACATT 2429
Qy 2101 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2160
Db 2430 GTGGACGTGCAGTACTTGTACGGGGTAGGGTCAAGCATCGCGTCCTGGGCCATTAAGTGG 2489
Qy 2161 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2220
Db 2490 GAGTACGTCGTTCTCCTGTTCCTCCTGCTTGCAGACGCGCGCGTCTGCTCCTGCTTGTGG 2549
Qy 2221 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2280
Db 2550 ATGATGTTACTCATATCCCAAGCGGAGGCGGCTTTGGAGAACCTCGTAATACTCAATGCA 2609
Qy 2281 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATATTTTGTCATCTTTTTCGTGGCTGCTTGG 2340
Db 2610 GCATCCCTGGCCGGGACGCACGGTCTTGTGTCCTTCCTCGTGTTCTTCTGCTTTGCGTGG 2669
Qy 2341 TACATCAAGGGTCGGGTAGTCCCCTTAGCTACCTATTCCCTCACTGGCCTGTGGTCCTTT 2400
Db 2670 TATCTGAAGGGTAGGTGGGTGCCCGGAGCGGTCTACGCCTTCTACGGGATGTGGCCTCTC 2729
Qy 2401 AGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTGCATGGC 2460
Db 2730 CTCCTGCTCCTGCTGGCGTTGCCTCAGCGGGCATACGCACTGGACACGGAGGTGGCCGCG 2789
Qy 2461 CAGATAGGAGCGGCTCTGCTGGTAATGATCACTCTCTTTACTCTCACCCCCGGGTATAAG 2520
Db 2790 TCGTGTGGCGGCGTTGTTCTTGTCGGGTTAATGGCGCTGACTCTGTCGCCATATTACAAG 2849
Qy 2521 ACCCTTCTCAGCCGGTTTTTGTGGTGGTTGTGCTATCTTCTGACCCTGGGGGAAGCTATG 2580
Db 2850 CGCTACATCAGCTGGTGCATGTGGTGGCTTCAGTATTTTCTGACCAGAGTAGAAGCGCAA 2909
Qy 2581 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGCATCATATGGGCC 2640
Db 2910 CTGCACGTGTGGGTTCCCCCCCTCAACGTCCGGGGGGGGCGCGATGCCGTCATCTTACTC 2969
Qy 2641 GTCGCCATATTCTACCCAGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 2700
Db 2970 ATGTGTGTTGTACACCCGACTCTGGTATTTGACATCACCAAACTACTCCTGGCCATCTTC 3029
Qy 2701 GGGCCTGCTTACCTCCTAAAAGGTGCTTTGACGCGCGTGCCGTACTTCGTCAGGGCTCAC 2760
Db 3030 GGACCCCTTTGGATTCTTCAAGCCAGTTTGCTTAAAGTCCCCTACTTCGTGCGCGTTCAA 3089
Qy 2761 GCTCTACTGAGGATGTGCACCATGGCAAGGCATCTCGCGGGGGGCAGGTACGTCCAGATG 2820
Db 3090 GGCCTTCTCCGGATCTGCGCGCTAGCGCGGAAGATAGCCGGAGGTCATTACGTGCAAATG 3149
Qy 2821 GCGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 2880
Db 3150 GCCATCATCAAGTTAGGGGCGCTTACTGGCACCTATGTGTATAACCATCTCACCCCTCTT 3209
Qy 2881 TCGGATTGGGCTGCTAGTGGCCTGCGGGACCTGGCGGTCGCCGTTGAGCCTATCATCTTC 2940
Db 3210 CGAGACTGGGCGCACAACGGCCTGCGAGATCTGGCCGTGGCTGTGGAACCAGTCGTCTTC 3269
Qy 2941 AGTCCGATGGAGAAGAAAGTCATTGTCTGGGGAGCGGAGACAGCTGCTTGTGGGGACATT 3000
Db 3270 TCCCGAATGGAGACCAAGCTCATCACGTGGGGGGCAGATACCGCCGCGTGCGGTGACATC 3329
Qy 3001 TTACACGGACTTCCCGTGTCCGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3060
Db 3330 ATCAACGGCTTGCCCGTCTCTGCCCGTAGGGGCCAGGAGATACTGCTTGGGCCAGCCGAC 3389
Qy 3061 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3120
Db 3390 GGAATGGTCTCCAAGGGGTGGAGGTTGCTGGCGCCCATCACGGCGTACGCCCAGCAGACG 3449
Qy 3121 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCT 3180
Db 3450 AGAGGCCTCCTAGGGTGTATAATCACCAGCCTGACTGGCCGGGACAAAAACCAAGTGGAG 3509
Qy 3181 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3240
Db 3510 GGTGAGGTCCAGATCGTGTCAACTGCTACCCAAACCTTCCTGGCAACGTGCATCAATGGG 3569
Qy 3241 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3300
Db 3570 GTATGCTGGACTGTCTACCACGGGGCCGGAACGAGGACCATCGCATCACCCAAGGGTCCT 3629
Qy 3301 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGG 3360
Db 3630 GTCATCCAGATGTATACCAATGTGGACCAAGACCTTGTGGGCTGGCCCGCTCCTCAAGGT 3689
Qy 3361 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3420
Db 3690 TCCCGCTCATTGACACCCTGCACCTGCGGCTCCTCGGACCTTTACCTGGTCACGAGGCAC 3749
Qy 3421 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3480
Db 3750 GCCGATGTCATTCCCGTGCGCCGGCGAGGTGATAGCAGGGGTAGCCTGCTTTCGCCCCGG 3809
Qy 3481 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3540
Db 3810 CCCATTTCCTACTTGAAAGGCTCCTCGGGGGGTCCGCTGTTGTGCCCCGCGGGACACGCC 3869
Qy 3541 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3600
Db 3870 GTGGGCCTATTCAGGGCCGCGGTGTGCACCCGTGGAGTGGCTAAGGCGGTGGACTTTATC 3929
Qy 3601 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 3660

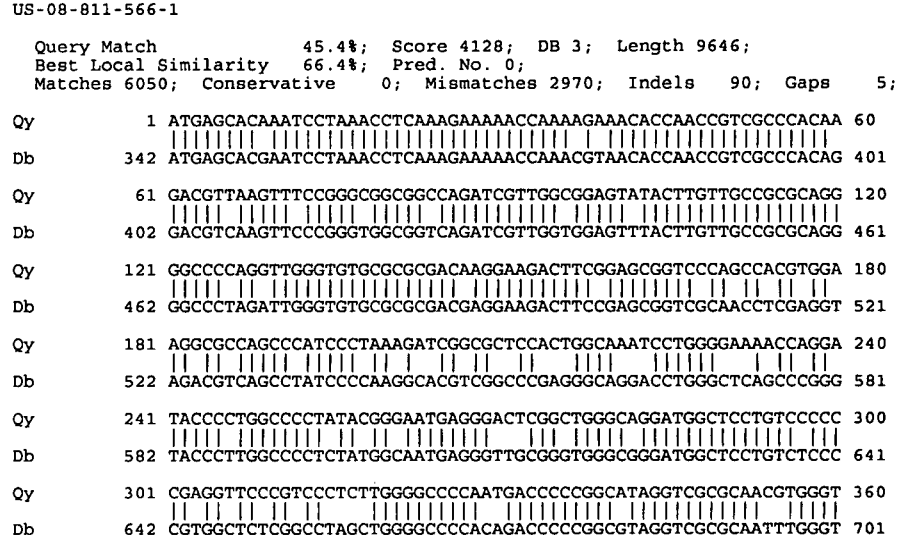

Db 3930 CCTGTGGAGAACCTAGAGACAACCATGAGATCCCCGGTGTTCACGGACAACTCCTCTCCA 3989
Qy 3661 CCTGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 3720
Db 3990 CCAGCAGTGCCCCAGAGCTTCCAGGTGGCCCACCTGCATGCTCCCACCGGCAGCGGTAAG 4049
Qy 3721 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 3780
Db 4050 AGCACCAAGGTCCCGGCTGCGTACGCAGCCCAGGGCTACAAGGTGTTGGTGCTCAACCCC 4109
Qy 3781 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 3840
Db 4110 TCTGTTGCTGCAACGCTGGGCTTTGGTGCTTACATGTCCAAGGCCCATGGGGTTGATCCT 4169
Qy 3841 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 3900
Db 4170 AATATCAGGACCGGGGTGAGAACAATTACCACTGGCAGCCCCATCACGTACTCCACCTAC 4229
Qy 3901 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 3960
Db 4230 GGCAAGTTCCTTGCCGACGGCGGGTGCTCAGGAGGTGCTTATGACATAATAATTTGTGAC 4289
Qy 3961 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4020
Db 4290 GAGTGCCACTCCACGGATGCCACATCCATCTTGGGCATCGGCACTGTCCTTGACCAAGCA 4349
Qy 4021 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4080
Db 4350 GAGACTGCGGGGGCGAGACTGGTTGTGCTCGCCACTGCTACCCCTCCGGGCTCCGTCACT 4409
Qy 4081 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4140
Db 4410 GTGTCCCATCCTAACATCGAGGAGGTTGCTCTGTCCACCACCGGAGAGATCCCCTTTTAC 4469
Qy 4141 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4200
Db 4470 GGCAAGGCTATCCCCCTCGAGGTGATCAAGGGGGGAAGACATCTCATCTTCTGCCACTCA 4529
Qy 4201 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4260
Db 4530 AAGAAGAAGTGCGACGAGCTCGCCGCGAAGCTGGTCGCATTGGGCATCAATGCCGTGGCC 4589
Qy 4261 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4320
Db 4590 TACTACCGCGGTCTTGACGTGTCTGTCATCCCGACCAGCGGCGATGTTGTCGTCGTGTCG 4649
Qy 4321 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4380
Db 4650 ACCGATGCTCTCATGACTGGCTTTACCGGCGACTTCGACTCTGTGATAGACTGCAACACG 4709
Qy 4381 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4440
Db 4710 TGTGTCACTCAGACAGTCGATTTCAGCCTTGACCCTACCTTTACCATTGAGACAACCACG 4769
Qy 4441 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4500
Db 4770 CTCCCCCAGGATGCTGTCTCCAGGACTCAACGCCGGGGCAGGACTGGCAGGGGGAAGCCA 4829
Qy 4501 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4560
Db 4830 GGCATCTACAGATTTGTGGCACCGGGGGAGCGCCCCTCCGGCATGTTCGACTCGTCCGTC 4889
Qy 4561 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4620
Db 4890 CTCTGTGAGTGCTATGACGCGGGCTGTGCTTGGTATGAGCTCACGCCCGCCGAGACTACA 4949
Qy 4621 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 4680
Db 4950 GTTAGGCTACGAGCGTACATGAACACCCCGGGGCTTCCCGTGTGCCAGGACCATCTTGAA 5009
Qy 4681 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 4740
Db 5010 TTTTGGGAGGGCGTCTTTACGGGCCTCACTCATATAGATGCCCACTTTCTATCCCAGACA 5069
Qy 4741 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 4800
Db 5070 AAGCAGAGTGGGGAGAACTTTCCTTACCTGGTAGCGTACCAAGCCACCGTGTGCGCTAGG 5129
Qy 4801 GCCAAAGCCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 4860
Db 5130 GCTCAAGCCCCTCCCCCATCGTGGGACCAGATGTGGAAGTGTTTGATCCGCCTTAAACCC 5189
Qy 4861 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 4920
Db 5190 ACCCTCCATGGGCCAACACCCCTGCTATACAGACTGGGCGCTGTTCAGAATGAAGTCACC 5249
Qy 4921 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 4980
Db 5250 CTGACGCACCCAATCACCAAATACATCATGACATGCATGTCGGCCGACCTGGAGGTCGTC 5309
Qy 4981 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5040
Db 5310 ACGAGCACCTGGGTGCTCGTTGGCGGCGTCCTGGCTGCTCTGGCCGCGTATTGCCTGTCA 5369
Qy 5041 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5100
Db 5370 ACAGGCTGCGTGGTCATAGTGGGCAGGATTGTCTTGTCCGGGAAGCCGGCAATTATACCT 5429
Qy 5101 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5160
Db 5430 GACAGGGAGGTTCTCTACCAGGAGTTCGATGAGATGGAAGAGTGCTCTCAGCACTTACCG 5489
Qy 5161 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5220
Db 5490 TACATCGAGCAAGGGATGATGCTCGCTGAGCAGTTCAAGCAGAAGGCCCTCGGCCTCCTG 5549
Qy 5221 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5280
Db 5550 CAGACCGCGTCCCGCCAAGCAGAGGTTATCACCCCTGCTGTCCAGACCAACTGGCAGAAA 5609
Qy 5281 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5340
Db 5610 CTCGAGGTCTTCTGGGCGAAGCACATGTGGAATTTCATCAGTGGGATACAATACTTGGCG 5669
Qy 5341 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5400
Db 5670 GGCCTGTCAACGCTGCCTGGTAACCCCGCCATTGCTTCATTGATGGCTTTTACAGCTGCC 5729
Qy 5401 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5460
Db 5730 GTCACCAGCCCACTAACCACTGGCCAAACCCTCCTCTTCAACATATTGGGGGGGTGGGTG 5789
Qy 5461 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5520
Db 5790 GCTGCCCAGCTCGCCGCCCCCGGTGCCGCTACCGCCTTTGTGGGCGCTGGCTTAGCTGGC 5849
Qy 5521 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5580
Db 5850 GCCGCCATCGGCAGCGTTGGACTGGGGAAGGTCCTCGTGGACATTCTTGCAGGGTATGGC 5909
Qy 5581 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5640
Db 5910 GCGGGCGTGGCGGGAGCTCTTGTAGCCTTCAAGATCATGAGCGGTGAGGTCCCCTCCACG 5969
Qy 5641 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 5700
Db 5970 GAGGACCTGGTCAATCTGCTGCCCGCCATCCTCTCGCCTGGAGCCCTTGTAGTCGGTGTG 6029
Qy 5701 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 5760
Db 6030 GTCTGCGCAGCAATACTGCGCCGGCACGTTGGCCCGGGCGAGGGGGCAGTGCAATGGATG 6089
Qy 5761 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 5820
Db 6090 AACCGGCTAATAGCCTTCGCCTCCCGGGGGAACCATGTTTCCCCCACGCACTACGTGCCG 6149

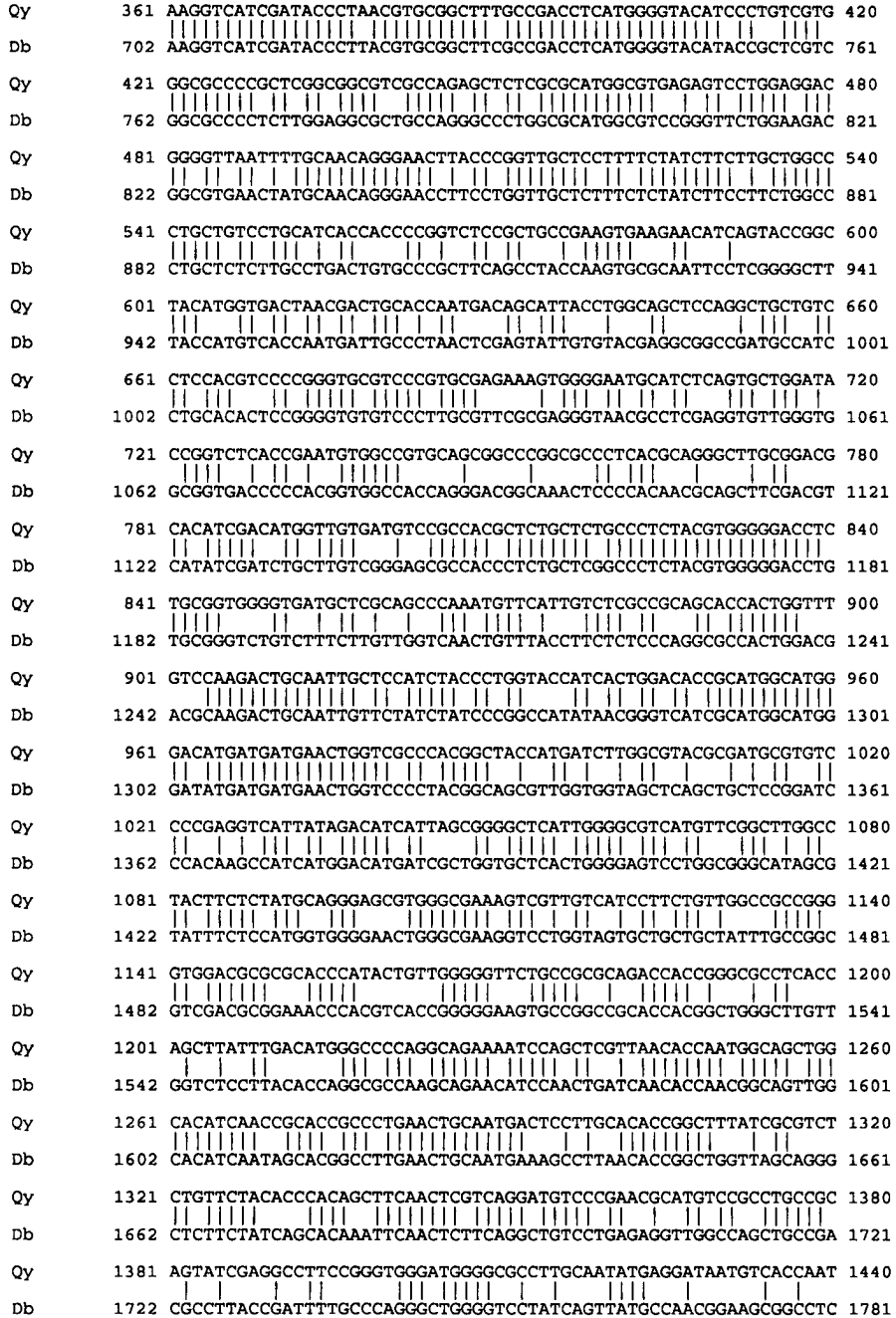

Qy 5821 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 5880
Db 6150 GAGAGCGATGCAGCCGCCCGCGTCACTGCCATACTCAGCAGCCTCACTGTAACCCAGCTC 6209
Qy 5881 CTCAGAAGACTCCACAACTGGATTACTGAGGACTGCCCCATCCCATGCGGCGGCTCGTGG 5940
Db 6210 CTGAGGCGACTGCATCAGTGGATAAGCTCGGAGTGTACCACTCCATGCTCCGGTTCCTGG 6269
Qy 5941 CTCCGCGATGTGTGGGACTGGGTTTGCACCATCCTAACAGACTTTAAAAATTGGCTGACC 6000
Db 6270 CTAAGGGACATCTGGGACTGGATATGCGAGGTGCTGAGCGACTTTAAGACCTGGCTGAAA 6329
Qy 6001 TCCAAATTATTCCCAAAGATGCCCGGCCTCCCCTTTGTCTCCTGTCAAAAGGGGTACAAG 6060
Db 6330 GCCAAGCTCATGCCACAACTGCCTGGGATTCCCTTTGTGTCCTGCCAGCGCGGGTATAGG 6389
Qy 6061 GGCGTGTGGGCCGGCACTGGCATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6120
Db 6390 GGGGTCTGGCGAGGAGACGGCATTATGCACACTCGCTGCCACTGTGGAGCTGAGATCACT 6449
Qy 6121 GGCAATGTCCGCTTGGGCTCCATGAGAATCACGGGGCCTAAGACCTGCATGAATATCTGG 6180
Db 6450 GGACATGTCAAAAACGGGACGATGAGGATCGTCGGTCCTAGGACCTGCAGGAACATGTGG 6509
Qy 6181 CAGGGGACCTTTCCTATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCGCCA 6240
Db 6510 AGTGGGACGTTCCCCATTAACGCCTACACCACGGGCCCCTGTACTCCCCTTCCTGCGCCG 6569
Qy 6241 AACTTTAAGGTCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6300
Db 6570 AACTATAAGTTCGCGCTGTGGAGGGTGTCTGCAGAGGAATACGTGGAGATAAGGCGGGTG 6629
Qy 6301 GGGTCATACCACTACATAACAGGACTCACCACTGATAACTTGAAAGTCCCCTGCCAACTA 6360
Db 6630 GGGGACTTCCACTACGTATCGGGTATGACTACTGACAATCTTAAATGCCCGTGCCAGATC 6689
Qy 6361 CCCTCTCCCGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCACA 6420
Db 6690 CCATCGCCCGAATTTTTCACAGAATTGGACGGGGTGCGCCTACATAGGTTTGCGCCCCCT 6749
Qy 6421 CCGAAGCCGTTTTTCCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6480
Db 6750 TGCAAGCCCTTGCTGCGGGAGGAGGTATCATTCAGAGTAGGACTCCACGAGTACCCGGTG 6809
Qy 6481 GGGTCCCAGCTTCCTTGCGACCCTGAACCCGACACAGACGTATTGATGTCCATGCTAACA 6540
Db 6810 GGGTCGCAATTACCTTGCGAGCCCGAACCGGACGTAGCCGTGTTGACGTCCATGCTCACT 6869
Qy 6541 GATCCATCTCATATCACGGCGGAGACTGCAGCGCGGCGTTTAGCGCGGGGGTCACCCCCA 6600
Db 6870 GATCCCTCCCATATAACAGCAGAGGCGGCCGGGAGAAGGTTGGCGAGAGGGTCACCCCCT 6929
Qy 6601 TCCGAGGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 6660
Db 6930 TCTATGGCCAGCTCCTCGGCCAGCCAGCTGTCCGCTCCATCTCTCAAGGCAACTTGCACC 6989
Qy 6661 ACCCACGGCAAAGCCTATGATGTGGACATGGTGGATGCTAACCT----------GTTC 6708
Db 6990 GCCAACCATGACTCCCCTGACGCCGAGCTCATAGAGGCTAACCTCCTGTGGAGGCAGGAG 7049
Qy 6709 ATGGGGGGCGATGTGACTCGGATAGAGTCTGGGTCCAAAGTGGTCGTTCTGGACTCTCTC 6768
Db 7050 ATGGGCGGCAACATCACCAGGGTTGAGTCAGAGAACAAAGTGGTGATTCTGGACTCCTTC 7109
Qy 6769 GACCCAATGGTCGAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCAGAATACATGCTC 6828
Db 7110 GATCCGCTTGTGGCAGAGGAGGATGAGCGGGAGGTCTCCGTACCCGCAGAAATTCTGCGG 7169
Qy 6829 CCCAAGAAGAGGTTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCG 6888
Db 7170 AAGTCTCGGAGATTCGCCCGGGCCCTGCCCGTTTGGGCGCGGCCGGACTACAACCCCCCG 7229
Qy 6889 CTTGTGGAATCGTGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTC 6948
Db 7230 CTAGTAGAGACGTGGAAAAAGCCTGACTACGAACCACCTGTGGTCCATGGCTGCCCGCTA 7289
Qy 6949 CCTCCTCCTAGGAAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGT 7008
Db 7290 CCACCTCCACGGTCCCCTCCTGTGCCTCCGCCTCGGAAAAAGCGTACGGTGGTCCTCACC 7349
Qy 7009 GAGGACTCCATAGGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCC 7068
Db 7350 GAATCAACCCTATCTACTGCCTTGGCCGAGCTTGCCACCAAAAGTTTTGGCAGCTCCTCA 7409
Qy 7069 CCAAGCGGCGATTCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACG 7128
Db 7410 ACTTCCGGCATTACGGGCGACAATACGACAACATCCTCTGAGCCCGCC---------- 7457
Qy 7129 CCTCCTGATGAGTTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCTCGAG 7188
Db 7458 CCTTCTGGCTGCCCCCCCGACTCCGACGTTGAGTCCTATTCTTCCATGCCCCCCCTGGAG 7517
Qy 7189 GGGGAGCTTGGAGATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCAG 7248
Db 7518 GGGGAGCCTGGGGATCCGGA---------- 7537
Qy 7249 GGGGGGGTGGCAGCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGAC 7308
Db 7538 ----------TCTCAGCGACGGGTCATGGTCGACGGTCAGTAGTGGGGCCGACACG 7583
Qy 7309 GACTCCGTCGTGTGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGT 7368
Db 7584 GAAGATGTCGTGTGCTGCTCAATGTCTTATTCCTGGACAGGCGCACTCGTCACCCCGTGC 7643
Qy 7369 AGTCCCGAAGAGGAGAAGTTACCGATTAACCCCTTGAGCAACTCCCTGTTGCGATATCAC 7428
Db 7644 GCTGCGGAAGAACAAAAACTGCCCATCAACGCACTGAGCAACTCGTTGCTACGCCATCAC 7703
Qy 7429 AACAAGGTGTACTGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTT 7488
Db 7704 AATCTGGTGTATTCCACCACTTCACGCAGTGCTTGCCAAAGGCAGAAGAAAGTCACATTT 7763
Qy 7489 GATAGGATGCAAGTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCG 7548
Db 7764 GACAGACTGCAAGTTCTGGACAGCCATTACCAGGACGTGCTCAAGGAGGTCAAAGCAGCG 7823
Qy 7549 GCCTCCAAGGTCACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCC 7608
Db 7824 GCGTCAAAAGTGAAGGCTAACTTGCTATCCGTAGAGGAAGCTTGCAGCCTGACGCCCCCA 7883
Qy 7609 CATTCTGCAAGATCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGG 7668
Db 7884 CATTCAGCCAAATCCAAGTTTGGCTATGGGGCAAAAGACGTCCGTTGCCATGCCAGAAAG 7943
Qy 7669 GCCGTTAACCACATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATT 7728
Db 7944 GCCGTAGCCCACATCAACTCCGTGTGGAAAGACCTTCTGGAAGACAGTGTAACACCAATA 8003
Qy 7729 CCCACAACCATTATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAG 7788
Db 8004 GACACTACCATCATGGCCAAGAACGAGGTTTTCTGCGTTCAGCCTGAGAAGGGGGGTCGT 8063
Qy 7789 AAAGCAGCTCGCCTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCC 7848
Db 8064 AAGCCAGCTCGTCTCATCGTGTTCCCCGACCTGGGCGTGCGCGTGTGCGAGAAGATGGCC 8123
Qy 7849 CTTTATGACATTACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAG 7908
Db 8124 CTGTACGACGTGGTTAGCAAGCTCCCCCTGGCCGTGATGGGAAGCTCCTACGGATTCCAA 8183
Qy 7909 TATTCCCCCGCTCAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCT 7968
Db 8184 TACTCACCAGGACAGCGGGTTGAATTCCTCGTGCAAGCGTGGAAGTCCAAGAAGACCCCG 8243
Qy 7969 ATGGGTTTTTCGTATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGG 8028

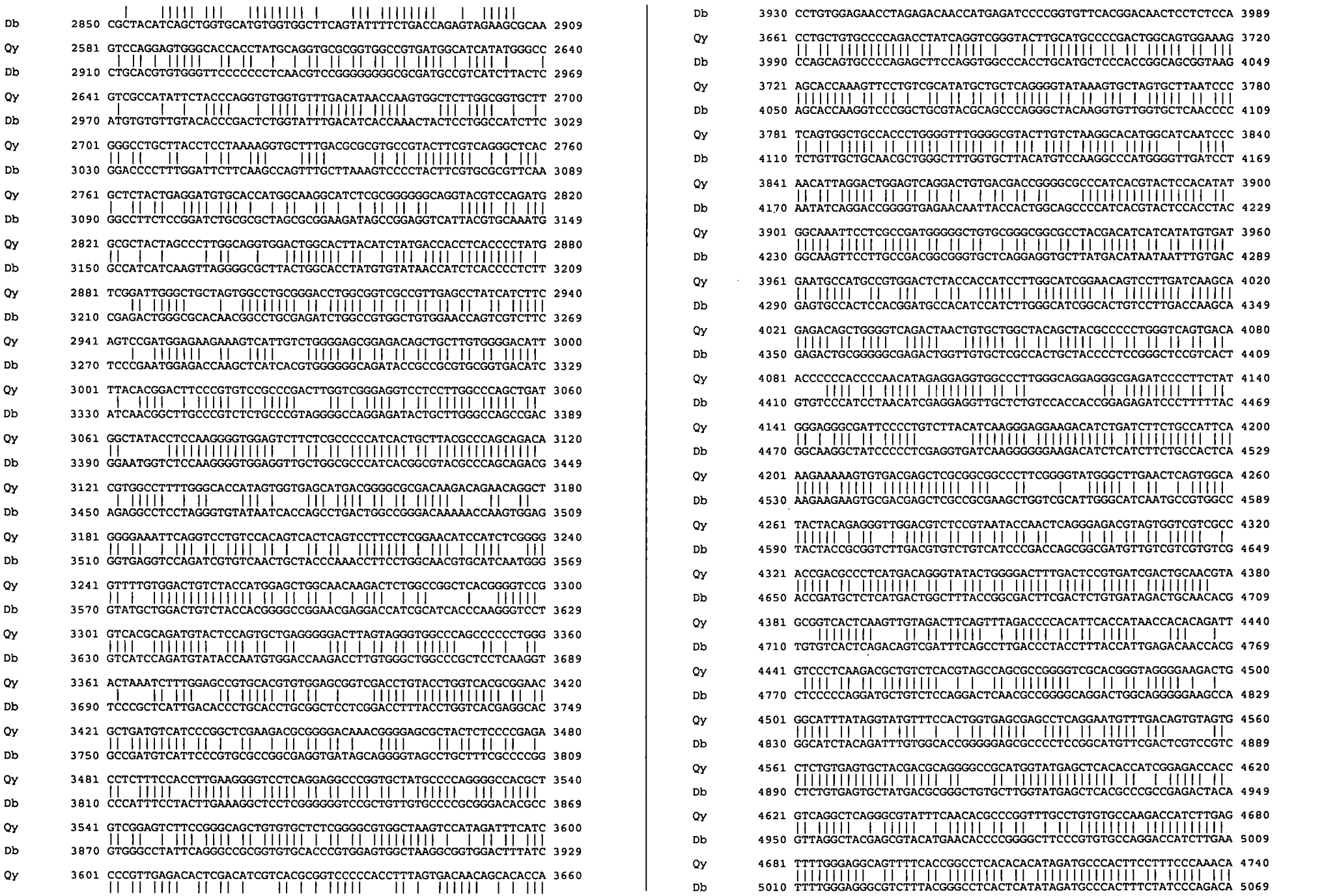


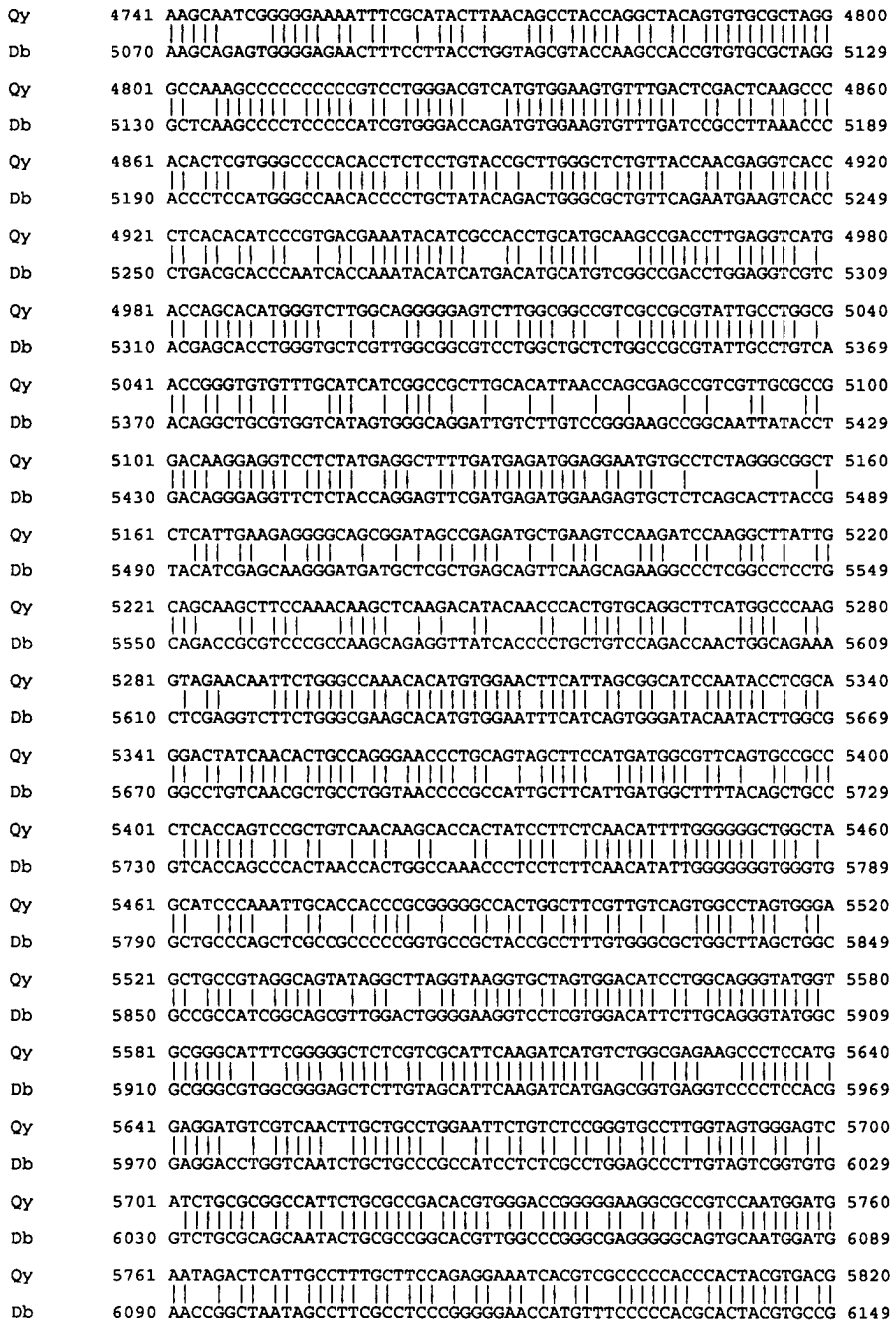
RESULT 9
US-08-811-566-1
; Sequence 1, Application US/08811566
; Patent No. 6127116
; GENERAL INFORMATION:
; APPLICANT: Rice, Charles et al.
; TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
; TITLE OF INVENTION: VIRUS (HCV) AND USES THEREOF
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; STREET: Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/811,566
; FILING DATE: 03-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 1113-1-006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9646 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-811-566-1

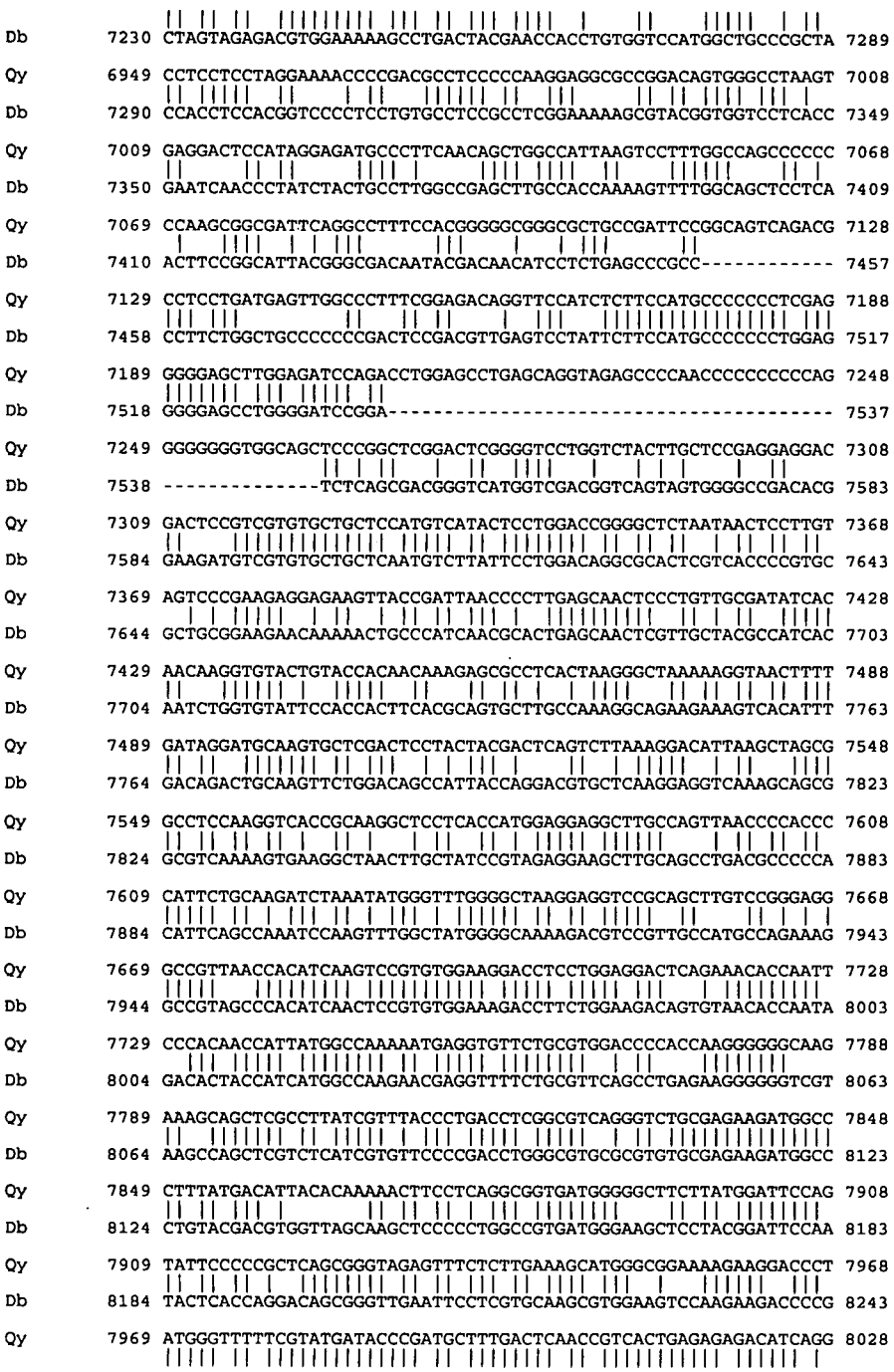
Query Match 45.4%; Score 4128; DB 3; Length 9646;
Best Local Similarity 66.4%; Pred. No. 0;
Matches 6050; Conservative 0; Mismatches 2970; Indels 90; Gaps 5;




Qy 361 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 420
Db 702 AAGGTCATCGATACCCTTACGTGCGGCTTCGCCGACCTCATGGGGTACATACCGCTCGTC 761
Qy 421 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 480
Db 762 GGCGCCCCTCTTGGAGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAAGAC 821
Qy 481 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 540
Db 822 GGCGTGAACTATGCAACAGGGAACCTTCCTGGTTGCTCTTTCTCTATCTTCCTTCTGGCC 881
Qy 541 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 600
Db 882 CTGCTCTCTTGCCTGACTGTGCCCGCTTCAGCCTACCAAGTGCGCAATTCCTCGGGGCTT 941
Qy 601 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 660
Db 942 TACCATGTCACCAATGATTGCCCTAACTCGAGTATTGTGTACGAGGCGGCCGATGCCATC 1001
Qy 661 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 720
Db 1002 CTGCACACTCCGGGGTGTGTCCCTTGCGTTCGCGAGGGTAACGCCTCGAGGTGTTGGGTG 1061
Qy 721 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 780
Db 1062 GCGGTGACCCCCACGGTGGCCACCAGGGACGGCAAACTCCCCACAACGCAGCTTCGACGT 1121
Qy 781 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 840
Db 1122 CATATCGATCTGCTTGTCGGGAGCGCCACCCTCTGCTCGGCCCTCTACGTGGGGGACCTG 1181
Qy 841 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 900
Db 1182 TGCGGGTCTGTCTTTCTTGTTGGTCAACTGTTTACCTTCTCTCCCAGGCGCCACTGGACG 1241
Qy 901 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 960
Db 1242 ACGCAAGACTGCAATTGTTCTATCTATCCCGGCCATATAACGGGTCATCGCATGGCATGG 1301
Qy 961 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1020
Db 1302 GATATGATGATGAACTGGTCCCCTACGGCAGCGTTGGTGGTAGCTCAGCTGCTCCGGATC 1361
Qy 1021 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1080
Db 1362 CCACAAGCCATCATGGACATGATCGCTGGTGCTCACTGGGGAGTCCTGGCGGGCATAGCG 1421
Qy 1081 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1140
Db 1422 TATTTCTCCATGGTGGGGAACTGGGCGAAGGTCCTGGTAGTGCTGCTGCTATTTGCCGGC 1481
Qy 1141 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1200
Db 1482 GTCGACGCGGAAACCCACGTCACCGGGGGAAGTGCCGGCCGCACCACGGCTGGGCTTGTT 1541
Qy 1201 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1260
Db 1542 GGTCTCCTTACACCAGGCGCCAAGCAGAACATCCAACTGATCAACACCAACGGCAGTTGG 1601
Qy 1261 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1320
Db 1602 CACATCAATAGCACGGCCTTGAACTGCAATGAAAGCCTTAACACCGGCTGGTTAGCAGGG 1661
Qy 1321 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1380
Db 1662 CTCTTCTATCAGCACAAATTCAACTCTTCAGGCTGTCCTGAGAGGTTGGCCAGCTGCCGA 1721
Qy 1381 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1440
Db 1722 CGCCTTACCGATTTTGCCCAGGGCTGGGGTCCTATCAGTTATGCCAACGGAAGCGGCCTC 1781
Qy 1441 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1500
Db 1782 GACGA------ACGCCCCTACTGCTGGCACTACCCTCCAAGACCTTGTGGCATTGTGCCC 1835
Qy 1501 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1560
Db 1836 GCAAAGAGCGTGTGTGGCCCGGTATATTGCTTCACTCCCAGCCCCGTGGTGGTGGGAACG 1895
Qy 1561 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1620
Db 1896 ACCGACAGGTCGGGCGCGCCTACCTACAGCTGGGGTGCAAATGATACGGATGTCTTCGTC 1955
Qy 1621 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 1680
Db 1956 CTTAACAACACCAGGCCACCGCTGGGCAATTGGTTCGGTTGTACCTGGATGAACTCAACT 2015
Qy 1681 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 1740
Db 2016 GGATTCACCAAAGTGTGCGGAGCGCCCCCTTGTGTCATCGGAGGGGTGGGCAACAACA-- 2073
Qy 1741 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 1800
Db 2074 ----CCTTGCTCTGCCCCACTGATTGTTTCCGCAAGCATCCGGAAGCCACATACTCTCGG 2129
Qy 1801 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 1860
Db 2130 TGCGGCTCCGGTCCCTGGATTACACCCAGGTGCATGGTCGACTACCCGTATAGGCTTTGG 2189
Qy 1861 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 1920
Db 2190 CACTATCCTTGTACCATCAATTACACCATATTCAAAGTCAGGATGTACGTGGGAGGGGTC 2249
Qy 1921 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 1980
Db 2250 GAGCACAGGCTGGAAGCGGCCTGCAACTGGACGCGGGGCGAACGCTGTGATCTGGAAGAC 2309
Qy 1981 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2040
Db 2310 AGGGACAGGTCCGAGCTCAGCCCATTGCTGCTGTCCACCACACAGTGGCAGGTCCTTCCG 2369
Qy 2041 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2100
Db 2370 TGTTCTTTCACGACCCTGCCAGCCTTGTCCACCGGCCTCATCCACCTCCACCAGAACATT 2429
Qy 2101 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2160
Db 2430 GTGGACGTGCAGTACTTGTACGGGGTAGGGTCAAGCATCGCGTCCTGGGCCATTAAGTGG 2489
Qy 2161 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2220
Db 2490 GAGTACGTCGTTCTCCTGTTCCTCCTGCTTGCAGACGCGCGCGTCTGCTCCTGCTTGTGG 2549
Qy 2221 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2280
Db 2550 ATGATGTTACTCATATCCCAAGCGGAGGCGGCTTTGGAGAACCTCGTAATACTCAATGCA 2609
Qy 2281 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATATTTTGTCATCTTTTTCGTGGCTGCTTGG 2340
Db 2610 GCATCCCTGGCCGGGACGCACGGTCTTGTGTCCTTCCTCGTGTTCTTCTGCTTTGCGTGG 2669
Qy 2341 TACATCAAGGGTCGGGTAGTCCCCTTAGCTACCTATTCCCTCACTGGCCTGTGGTCCTTT 2400
Db 2670 TATCTGAAGGGTAGGTGGGTGCCCGGAGCGGTCTACGCCTTCTACGGGATGTGGCCTCTC 2729
Qy 2401 AGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTGCATGGC 2460
Db 2730 CTCCTGCTCCTGCTGGCGTTGCCTCAGCGGGCATACGCACTGGACACGGAGGTGGCCGCG 2789
Qy 2461 CAGATAGGAGCGGCTCTGCTGGTAATGATCACTCTCTTTACTCTCACCCCCGGGTATAAG 2520
Db 2790 TCGTGTGGCGGCGTTGTTCTTGTCGGGTTAATGGCGCTGACTCTGTCGCCATATTACAAG 2849
Qy 2521 ACCCTTCTCAGCCGGTTTTTGTGGTGGTTGTGCTATCTTCTGACCCTGGGGGAAGCTATG 2580




Qy 4741 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 4800
Db 5070 AAGCAGAGTGGGGAGAACTTTCCTTACCTGGTAGCGTACCAAGCCACCGTGTGCGCTAGG 5129
Qy 4801 GCCAAAGCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 4860
Db 5130 GCTCAAGCCCCTCCCCCATCGTGGGACCAGATGTGGAAGTGTTTGATCCGCCTTAAACCC 5189
Qy 4861 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 4920
Db 5190 ACCCTCCATGGGCCAACACCCCTGCTATACAGACTGGGCGCTGTTCAGAATGAAGTCACC 5249
Qy 4921 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 4980
Db 5250 CTGACGCACCCAATCACCAAATACATCATGACATGCATGTCGGCCGACCTGGAGGTCGTC 5309
Qy 4981 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5040
Db 5310 ACGAGCACCTGGGTGCTCGTTGGCGGCGTCCTGGCTGCTCTGGCCGCGTATTGCCTGTCA 5369
Qy 5041 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5100
Db 5370 ACAGGCTGCGTGGTCATAGTGGGCAGGATTGTCTTGTCCGGGAAGCCGGCAATTATACCT 5429
Qy 5101 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5160
Db 5430 GACAGGGAGGTTCTCTACCAGGAGTTCGATGAGATGGAAGAGTGCTCTCAGCACTTACCG 5489
Qy 5161 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5220
Db 5490 TACATCGAGCAAGGGATGATGCTCGCTGAGCAGTTCAAGCAGAAGGCCCTCGGCCTCCTG 5549
Qy 5221 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5280
Db 5550 CAGACCGCGTCCCGCCAAGCAGAGGTTATCACCCCTGCTGTCCAGACCAACTGGCAGAAA 5609
Qy 5281 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5340
Db 5610 CTCGAGGTCTTCTGGGCGAAGCACATGTGGAATTTCATCAGTGGGATACAATACTTGGCG 5669
Qy 5341 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5400
Db 5670 GGCCTGTCAACGCTGCCTGGTAACCCCGCCATTGCTTCATTGATGGCTTTTACAGCTGCC 5729
Qy 5401 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5460
Db 5730 GTCACCAGCCCACTAACCACTGGCCAAACCCTCCTCTTCAACATATTGGGGGGGTGGGTG 5789
Qy 5461 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5520
Db 5790 GCTGCCCAGCTCGCCGCCCCCGGTGCCGCTACCGCCTTTGTGGGCGCTGGCTTAGCTGGC 5849
Qy 5521 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5580
Db 5850 GCCGCCATCGGCAGCGTTGGACTGGGGAAGGTCCTCGTGGACATTCTTGCAGGGTATGGC 5909
Qy 5581 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5640
Db 5910 GCGGGCGTGGCGGGAGCTCTTGTAGCATTCAAGATCATGAGCGGTGAGGTCCCCTCCACG 5969
Qy 5641 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 5700
Db 5970 GAGGACCTGGTCAATCTGCTGCCCGCCATCCTCTCGCCTGGAGCCCTTGTAGTCGGTGTG 6029
Qy 5701 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 5760
Db 6030 GTCTGCGCAGCAATACTGCGCCGGCACGTTGGCCCGGGCGAGGGGGCAGTGCAATGGATG 6089
Qy 5761 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 5820
Db 6090 AACCGGCTAATAGCCTTCGCCTCCCGGGGGAACCATGTTTCCCCCACGCACTACGTGCCG 6149
Qy 5821 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 5880
Db 6150 GAGAGCGATGCAGCCGCCCGCGTCACTGCCATACTCAGCAGCCTCACTGTAACCCAGCTC 6209
Qy 5881 CTCAGAAGACTCCACAACTGGATTACTGAGGACTGCCCCATCCCATGCGGCGGCTCGTGG 5940
Db 6210 CTGAGGCGACTGCATCAGTGGATAAGCTCGGAGTGTACCACTCCATGCTCCGGTTCCTGG 6269
Qy 5941 CTCCGCGATGTGTGGGACTGGGTTTGCACCATCCTAACAGACTTTAAAAAATTGGCTGACC 6000
Db 6270 CTAAGGGACATCTGGGACTGGATATGCGAGGTGCTGAGCGACTTTAAGACCTGGCTGAAA 6329
Qy 6001 TCCAAATTATTCCCAAAGATGCCCGGCCTCCCCTTTGTCTCCTGTCAAAAGGGGTACAAG 6060
Db 6330 GCCAAGCTCATGCCACAACTGCCTGGGATTCCCTTTGTGTCCTGCCAGCGCGGGTATAGG 6389
Qy 6061 GGCGTGTGGGCCGGCACTGGCATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6120
Db 6390 GGGGTCTGGCGAGGAGACGGCATTATGCACACTCGCTGCCACTGTGGAGCTGAGATCACT 6449
Qy 6121 GGCAATGTCCGCTTGGGCTCCATGAGAATCACGGGGCCTAAGACCTGCATGAATATCTGG 6180
Db 6450 GGACATGTCAAAAACGGGACGATGAGGATCGTCGGTCCTAGGACCTGCAGGAACATGTGG 6509
Qy 6181 CAGGGGACCTTTCCTATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCGCCA 6240
Db 6510 AGTGGGACGTTCCCCATTAACGCCTACACCACGGGCCCCTGTACTCCCCTTCCTGCGCCG 6569
Qy 6241 AACTTTAAGGTCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6300
Db 6570 AACTATAAGTTCGCGCTGTGGAGGGTGTCTGCAGAGGAATACGTGGAGATAAGGCGGGTG 6629
Qy 6301 GGGTCATACCACTACATAACAGGACTCACCACTGATAACTTGAAAGTCCCCTGCCAACTA 6360
Db 6630 GGGGACTTCCACTACGTATCGGGTATGACTACTGACAATCTTAAATGCCCGTGCCAGATC 6689
Qy 6361 CCCTCTCCCGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCACA 6420
Db 6690 CCATCGCCCGAATTTTTCACAGAATTGGACGGGGTGCGCCTACATAGGTTTGCGCCCCCT 6749
Qy 6421 CCGAAGCCGTTTTTCCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6480
Db 6750 TGCAAGCCCTTGCTGCGGGAGGAGGTATCATTCAGAGTAGGACTCCACGAGTACCCGGTG 6809
Qy 6481 GGGTCCCAGCTTCCTTGCGACCCTGAACCCGACACAGACGTATTGATGTCCATGCTAACA 6540
Db 6810 GGGTCGCAATTACCTTGCGAGCCCGAACCGGACGTAGCCGTGTTGACGTCCATGCTCACT 6869
Qy 6541 GATCCATCTCATATCACGGCGGAGACTGCAGCGCGGCGTTTAGCGCGGGGGTCACCCCCA 6600
Db 6870 GATCCCTCCCATATAACAGCAGAGGCGGCCGGGAGAAGGTTGGCGAGAGGGTCACCCCCT 6929
Qy 6601 TCCGAGGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 6660
Db 6930 TCTATGGCCAGCTCCTCGGCCAGCCAGCTGTCCGCTCCATCTCTCAAGGCAACTTGCACC 6989
Qy 6661 ACCCACGGCAAAGCCTATGATGTGGACATGGTGGATGCTAACCT----------GTTC 6708
Db 6990 GCCAACCATGACTCCCCTGACGCCGAGCTCATAGAGGCTAACCTCCTGTGGAGGCAGGAG 7049
Qy 6709 ATGGGGGGCGATGTGACTCGGATAGAGTCTGGGTCCAAAGTGGTCGTTCTGGACTCTCTC 6768
Db 7050 ATGGGCGGCAACATCACCAGGGTTGAGTCAGAGAACAAAGTGGTGATTCTGGACTCCTTC 7109
Qy 6769 GACCCAATGGTCGAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCAGAATACATGCTC 6828
Db 7110 GATCCGCTTGTGGCAGAGGAGGATGAGCGGGAGGTCTCCGTACCCGCAGAAATTCTGCGG 7169
Qy 6829 CCCAAGAAGAGGTTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCG 6888
Db 7170 AAGTCTCGGAGATTCGCCCGGGCCCTGCCCGTTTGGGCGCGGCCGGACTACAACCCCCCG 7229
Qy 6889 CTTGTGGAATCGTGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTC 6948


Db 7230 CTAGTAGAGACGTGGAAAAAGCCTGACTACGAACCACCTGTGGTCCATGGCTGCCCGCTA 7289
Qy 6949 CCTCCTCCTAGGAAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGT 7008
Db 7290 CCACCTCCACGGTCCCCTCCTGTGCCTCCGCCTCGGAAAAAGCGTACGGTGGTCCTCACC 7349
Qy 7009 GAGGACTCCATAGGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCC 7068
Db 7350 GAATCAACCCTATCTACTGCCTTGGCCGAGCTTGCCACCAAAAGTTTTGGCAGCTCCTCA 7409
Qy 7069 CCAAGCGGCGATTCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACG 7128
Db 7410 ACTTCCGGCATTACGGGCGACAATACGACAACATCCTCTGAGCCCGCC---------- 7457
Qy 7129 CCTCCTGATGAGTTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCTCGAG 7188
Db 7458 CCTTCTGGCTGCCCCCCCGACTCCGACGTTGAGTCCTATTCTTCCATGCCCCCCCTGGAG 7517
Qy 7189 GGGGAGCTTGGAGATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCAG 7248
Db 7518 GGGGAGCCTGGGGATCCGGA---------- 7537
Qy 7249 GGGGGGGTGGCAGCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGAC 7308
Db 7538 ----------TCTCAGCGACGGGTCATGGTCGACGGTCAGTAGTGGGGCCGACACG 7583
Qy 7309 GACTCCGTCGTGTGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGT 7368
Db 7584 GAAGATGTCGTGTGCTGCTCAATGTCTTATTCCTGGACAGGCGCACTCGTCACCCCGTGC 7643
Qy 7369 AGTCCCGAAGAGGAGAAGTTACCGATTAACCCCTTGAGCAACTCCCTGTTGCGATATCAC 7428
Db 7644 GCTGCGGAAGAACAAAAACTGCCCATCAACGCACTGAGCAACTCGTTGCTACGCCATCAC 7703
Qy 7429 AACAAGGTGTACTGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTT 7488
Db 7704 AATCTGGTGTATTCCACCACTTCACGCAGTGCTTGCCAAAGGCAGAAGAAAGTCACATTT 7763
Qy 7489 GATAGGATGCAAGTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCG 7548
Db 7764 GACAGACTGCAAGTTCTGGACAGCCATTACCAGGACGTGCTCAAGGAGGTCAAAGCAGCG 7823
Qy 7549 GCCTCCAAGGTCACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCC 7608
Db 7824 GCGTCAAAAGTGAAGGCTAACTTGCTATCCGTAGAGGAAGCTTGCAGCCTGACGCCCCCA 7883
Qy 7609 CATTCTGCAAGATCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGG 7668
Db 7884 CATTCAGCCAAATCCAAGTTTGGCTATGGGGCAAAAGACGTCCGTTGCCATGCCAGAAAG 7943
Qy 7669 GCCGTTAACCACATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATT 7728
Db 7944 GCCGTAGCCCACATCAACTCCGTGTGGAAAGACCTTCTGGAAGACAGTGTAACACCAATA 8003
Qy 7729 CCCACAACCATTATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAG 7788
Db 8004 GACACTACCATCATGGCCAAGAACGAGGTTTTCTGCGTTCAGCCTGAGAAGGGGGGTCGT 8063
Qy 7789 AAAGCAGCTCGCCTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCC 7848
Db 8064 AAGCCAGCTCGTCTCATCGTGTTCCCCGACCTGGGCGTGCGCGTGTGCGAGAAGATGGCC 8123
Qy 7849 CTTTATGACATTACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAG 7908
Db 8124 CTGTACGACGTGGTTAGCAAGCTCCCCCTGGCCGTGATGGGAAGCTCCTACGGATTCCAA 8183
Qy 7909 TATTCCCCCGCTCAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCT 7968
Db 8184 TACTCACCAGGACAGCGGGTTGAATTCCTCGTGCAAGCGTGGAAGTCCAAGAAGACCCCG 8243
Qy 7969 ATGGGTTTTTCGTATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGG 8028
Db 8244 ATGGGGTTCTCGTATGATACCCGCTGTTTTGACTCCACAGTCACTGAGAGCGACATCCGT 8303
Qy 8029 ACTGAGGAGTCCATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACAC 8088
Db 8304 ACGGAGGAGGCAATTTACCAATGTTGTGACCTGGACCCCCAAGCCCGCGTGGCCATCAAG 8363
Qy 8089 TCGCTAACTGAGAGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGC 8148
Db 8364 TCCCTCACTGAGAGGCTTTATGTTGGGGGCCCTCTTACCAATTCAAGGGGGGAAAACTGC 8423
Qy 8149 GGGTACAGGCGTTGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACA 8208
Db 8424 GGCTACCGCAGGTGCCGCGCGAGCGGCGTACTGACAACTAGCTGTGGTAACACCCTCACT 8483
Qy 8209 TGCTACGTGAAAGCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTG 8268
Db 8484 TGCTACATCAAGGCCCGGGCAGCCTGTCGAGCCGCAGGGCTCCAGGACTGCACCATGCTC 8543
Qy 8269 GTATGCGGCGATGACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGG 8328
Db 8544 GTGTGTGGCGACGACTTAGTCGTTATCTGTGAAAGTGCGGGGGTCCAGGAGGACGCGGCG 8603
Qy 8329 AACCTGAGAGCCTTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCC 8388
Db 8604 AGCCTGAGAGCCTTCACGGAGGCTATGACCAGGTACTCCGCCCCCCCCGGGGACCCCCCA 8663
Qy 8389 AGACCGGAGTATGATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTG 8448
Db 8664 CAACCAGAATACGACTTGGAGCTTATAACATCATGCTCCTCCAACGTGTCAGTCGCCCAC 8723
Qy 8449 GGCCCACAAGGCCGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGG 8508
Db 8724 GACGGCGCTGGAAAGAGGGTCTACTACCTTACCCGTGACCCTACAACCCCCCTCGCGAGA 8783
Qy 8509 GCTGCCTGGGAAACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAG 8568
Db 8784 GCCGCGTGGGAGACAGCAAGACACACTCCAGTCAATTCCTGGCTAGGCAACATAATCATG 8843
Qy 8569 TACGCCCCGACCATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATG 8628
Db 8844 TTTGCCCCCACACTGTGGGCGAGGATGATACTGATGACCCATTTCTTTAGCGTCCTCATA 8903
Qy 8629 GCTCAAGACACGCTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTG 8688
Db 8904 GCCAGGGATCAGCTTGAACAGGCTCTTAACTGTGAGATCTACGGAGCCTGCTACTCCATA 8963
Qy 8689 AGTCCCTTGGACCTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTG 8748
Db 8964 GAACCACTGGATCTACCTCCAATCATTCAAAGACTCCATGGCCTCAGCGCATTTTCACTC 9023
Qy 8749 CACACATACACTCCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCG 8808
Db 9024 CACAGTTACTCTCCAGGTGAAATCAATAGGGTGGCCGCATGCCTCAGAAAACTTGGGGTC 9083
Qy 8809 CCACCCCTCAGAGCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGT 8868
Db 9084 CCGCCCTTGCGAGCTTGGAGACACCGGGCCCGGAGCGTCCGCGCTAGGCTTCTGTCCAGA 9143
Qy 8869 GGGGGGAGAGCGGCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTC 8928
Db 9144 GGAGGCAGGGCTGCCATATGTGGCAAGTACCTCTTCAACTGGGCAGTAAGAACAAAGCTC 9203
Qy 8929 AAACTCACTCCATTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGC 8988
Db 9204 AAACTCACTCCAATAGCGGCCGCTGGCCGGCTGGACTTGTCCGGTTGGTTCACGGCTGGC 9263
Qy 8989 GCCGGCGGGGGCGACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTT 9048
Db 9264 TACAGCGGGGGAGACATTTATCACAGCGTGTCTCATGCCCGGCCCCGCTGGTTCTGGTTT 9323
Qy 9049 GGCCTACTCCTACTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCG 9098
Db 9324 TGCCTACTCCTGCTCGCTGCAGGGGTAGGCATCTACCTCCTCCCCAACCG 9373

```
RESULT 10
US-09-034-756-1
; Sequence 1, Application US/09034756
; Patent No. 6392028
;   GENERAL INFORMATION:
;        APPLICANT: RICE, CHARLES  et al.
;        TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
;                            VIRUS (HCV) AND USES THEREOF
;        NUMBER OF SEQUENCES: 21
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: HOWELL & HAFERKAMP, L.C.
;             STREET: 7733 FORSYTH BLVD., SUITE 1400
;             CITY: ST. LOUIS
;             STATE: MO
;             COUNTRY: USA
;             ZIP: 63105
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/034,756
;             FILING DATE: 04-May-1998
;             CLASSIFICATION: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: HOLLAND, DONALD R.
;             REGISTRATION NUMBER: 35,197
;             REFERENCE/DOCKET NUMBER: 6029-4831
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 314-727-5188
;             TELEFAX: 314-727-6092
;   INFORMATION FOR SEQ ID NO: 1:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 9646 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        MOLECULE TYPE: cDNA
;        HYPOTHETICAL: NO
;        ANTI-SENSE: NO
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-034-756-1

  Query Match             45.4%;  Score 4128;  DB 3;  Length 9646;
  Best Local Similarity   66.4%;  Pred. No. 0;
  Matches 6050;  Conservative    0;  Mismatches 2970;  Indels   90;  Gaps    5;

Qy          1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 60
              |||||||| |||||||||||||||||||||||||||| | ||||||||||||||||||||
Db        342 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGTCGCCCACAG 401

Qy         61 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 120
              ||||| ||||| ||||| ||||| |||||||||| ||||| ||||||||||||||||||
Db        402 GACGTCAAGTTCCCGGGTGGCGGTCAGATCGTTGGTGGAGTTTACTTGTTGCCGCGCAGG 461

Qy        121 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 180
              ||||| || |||||||||||||||||| |||||||||| |||||||| || || || ||
Db        462 GGCCCTAGATTGGGTGTGCGCGCGACGAGGAAGACTTCCGAGCGGTCGCAACCTCGAGGT 521

Qy        181 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 240
              || || ||||| ||||| || |  || ||  ||  ||||  ||||||  | || ||
Db        522 AGACGTCAGCCTATCCCCAAGGCACGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGG 581

Qy        241 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 300
              ||||| |||||||| || || ||||||||   ||| ||||| ||||||||||||||| |||
Db        582 TACCCTTGGCCCCTCTATGGCAATGAGGGTTGCGGGTGGGCGGGATGGCTCCTGTCTCCC 641

Qy        301 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 360
              || || || || ||    |||||||||| |||||||||| ||||||||||||  |||||
Db        642 CGTGGCTCTCGGCCTAGCTGGGGCCCCACAGACCCCCGGCGTAGGTCGCGCAATTTGGGT 701

Qy        361 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 420
              |||||||||||||||| |||||||||| |||||||||||||||||||| ||  ||||
Db        702 AAGGTCATCGATACCCTTACGTGCGGCTTCGCCGACCTCATGGGGTACATACCGCTCGTC 761

Qy        421 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 480
              ||||||| || || |||| ||||| || || |||||||||| | || ||||| |||
Db        762 GGCGCCCCTCTTGGAGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAAGAC 821

Qy        481 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 540
              || || || | |||||||||||| | || |||||||| || |||||||| | ||||||
Db        822 GGCGTGAACTATGCAACAGGGAACCTTCCTGGTTGCTCTTTCTCTATCTTCCTTCTGGCC 881

Qy        541 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 600
              ||||| || ||| | ||    || |  || ||   | |||||   ||   |
Db        882 CTGCTCTCTTGCCTGACTGTGCCCGCTTCAGCCTACCAAGTGCGCAATTCCTCGGGGCTT 941

Qy        601 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 660
              |||   || || || || ||| | ||    || |||   |   ||      | |||  ||
Db        942 TACCATGTCACCAATGATTGCCCTAACTCGAGTATTGTGTACGAGGCGGCCGATGCCATC 1001

Qy        661 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 720
              || |||   || ||||| ||||| ||||      | ||| || || ||  ||| ||| |
Db       1002 CTGCACACTCCGGGGTGTGTCCCTTGCGTTCGCGAGGGTAACGCCTCGAGGTGTTGGGTG 1061

Qy        721 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 780
              ||||  | || |  ||||||      |     |    || |||   |    | ||
Db       1062 GCGGTGACCCCCACGGTGGCCACCAGGGACGGCAAACTCCCCACAACGCAGCTTCGACGT 1121

Qy        781 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 840
              || |||||  || ||||    | |||||| |||||| |||||||||||||||||||||
Db       1122 CATATCGATCTGCTTGTCGGGAGCGCCACCCTCTGCTCGGCCCTCTACGTGGGGGACCTG 1181

Qy        841 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 900
              |||||   ||  | || |  |  ||| |||| |  |||| ||   || |||||||
Db       1182 TGCGGGTCTGTCTTTCTTGTTGGTCAACTGTTTACCTTCTCTCCCAGGCGCCACTGGACG 1241

Qy        901 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 960
              |||||||||||||| || ||||| || ||    || || || || ||||||||||||
Db       1242 ACGCAAGACTGCAATTGTTCTATCTATCCCGGCCATATAACGGGTCATCGCATGGCATGG 1301

Qy        961 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1020
              || |||||||||||||||| || |||||  |  ||| |  || |  | || ||
Db       1302 GATATGATGATGAACTGGTCCCCTACGGCAGCGTTGGTGGTAGCTCAGCTGCTCCGGATC 1361

Qy       1021 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1080
              ||  | | ||| || ||||| ||     || ||||| ||||| ||| ||   ||| | ||
Db       1362 CCACAAGCCATCATGGACATGATCGCTGGTGCTCACTGGGGAGTCCTGGCGGGCATAGCG 1421

Qy       1081 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1140
              || ||||| ||| |||    |||||||| ||| | ||  | || ||| |    |||||
Db       1422 TATTTCTCCATGGTGGGGAACTGGGCGAAGGTCCTGGTAGTGCTGCTGCTATTTGCCGGC 1481

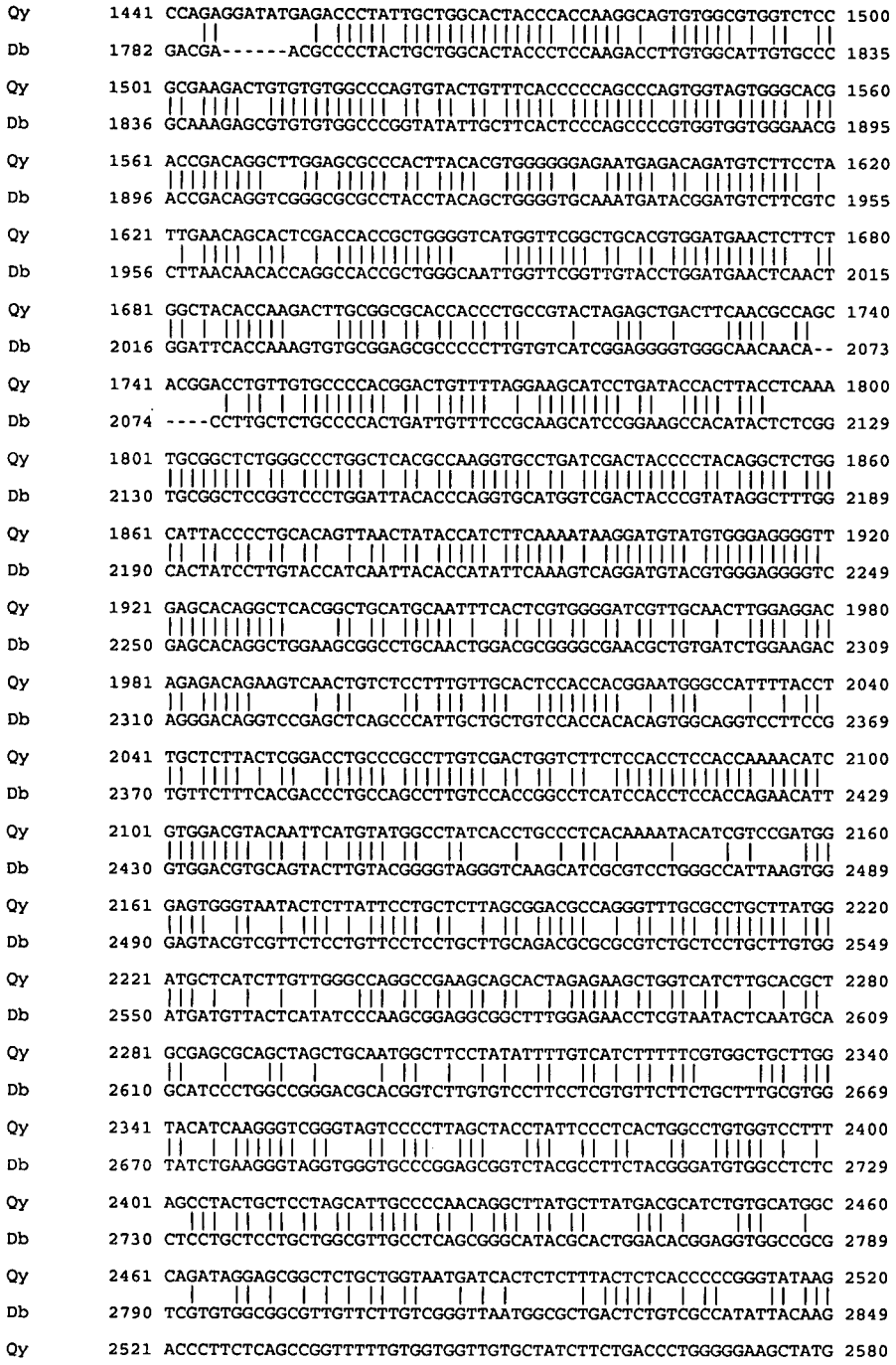
Qy       1141 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1200
              || ||||||    |||||         |||||   |||||  |   | ||
Db       1482 GTCGACGCGGAAACCCACGTCACCGGGGGAAGTGCCGGCCGCACCACGGCTGGGCTTGTT 1541

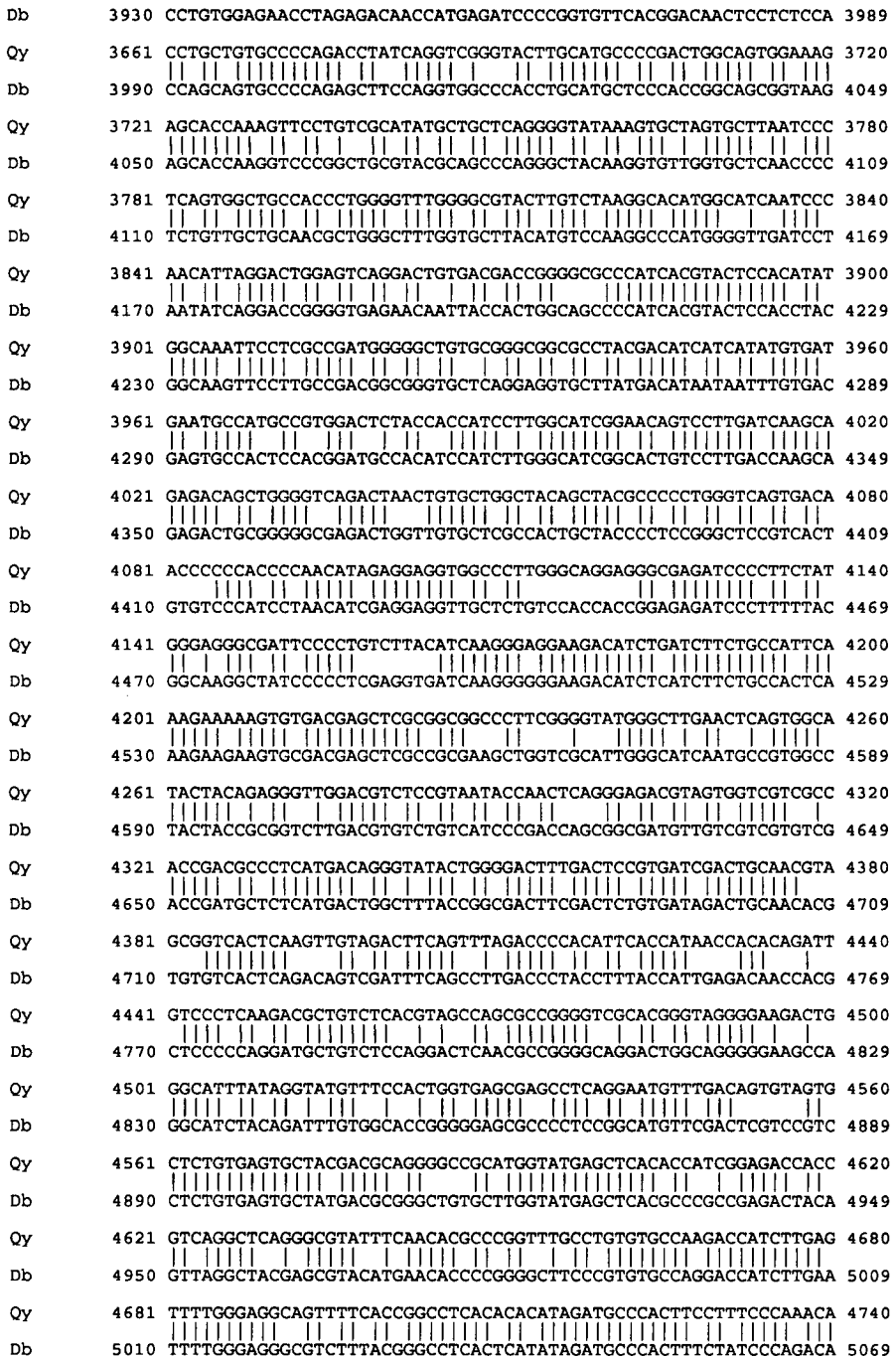
Qy       1201 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1260
              |  |  ||     ||| ||| ||||| ||||| ||  | |||||||| |||||  |||
Db       1542 GGTCTCCTTACACCAGGCGCCAAGCAGAACATCCAACTGATCAACACCAACGGCAGTTGG 1601

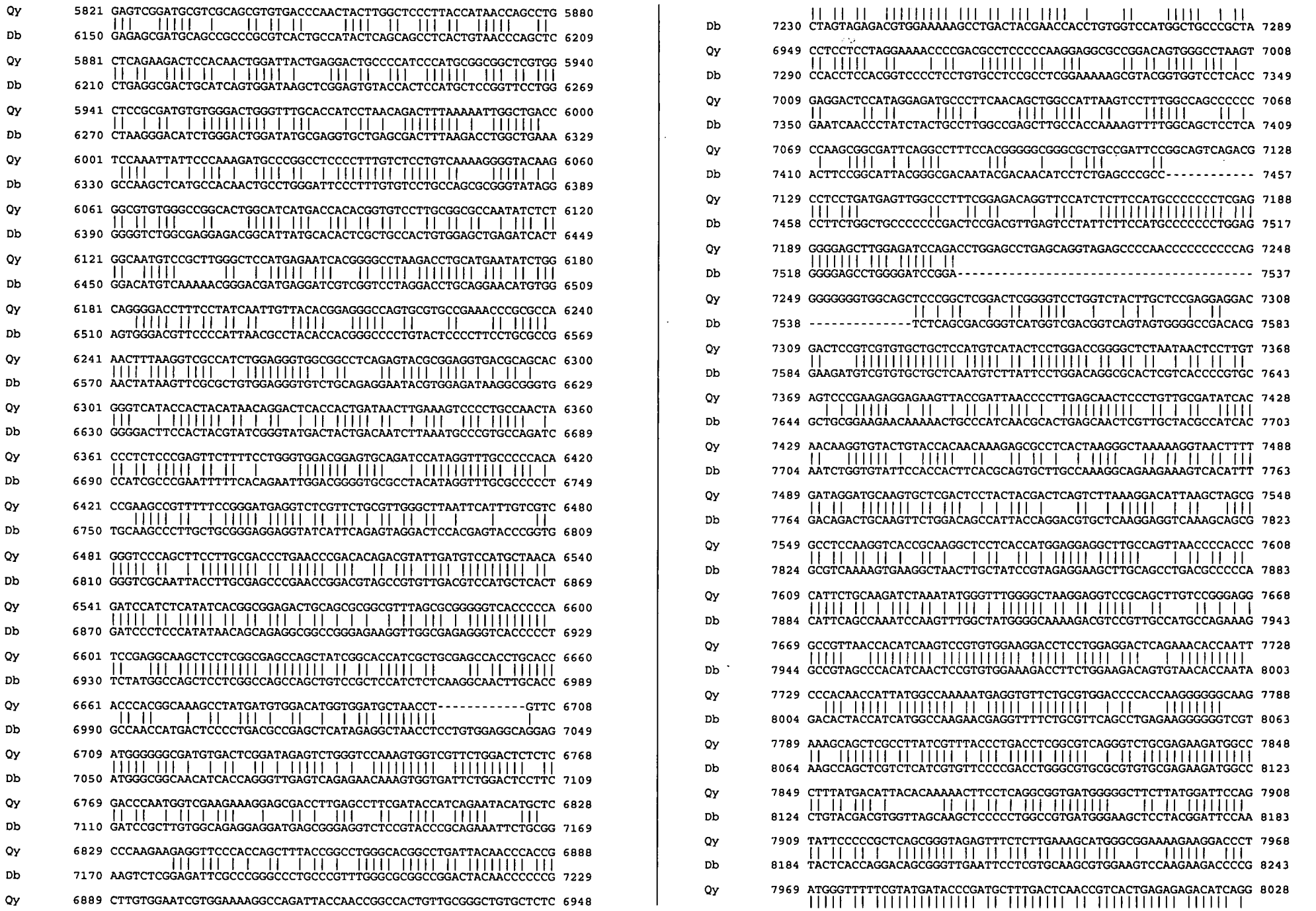
Qy       1261 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1320
              ||||||||  |||| ||| |||||||||||||   | |  |||||||||   | ||
Db       1602 CACATCAATAGCACGGCCTTGAACTGCAATGAAAGCCTTAACACCGGCTGGTTAGCAGGG 1661

Qy       1321 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1380
              || |||||     |||| |||||||| ||||| ||||| ||  |  || ||  ||||||
Db       1662 CTCTTCTATCAGCACAAATTCAACTCTTCAGGCTGTCCTGAGAGGTTGGCCAGCTGCCGA 1721

Qy       1381 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1440
              |  |  |  ||     ||| |||||  |  |   ||||  |   |  |
Db       1722 CGCCTTACCGATTTTGCCCAGGGCTGGGGTCCTATCAGTTATGCCAACGGAAGCGGCCTC 1781
```


Qy 1441 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1500
Db 1782 GACGA------ACGCCCCTACTGCTGGCACTACCCTCCAAGACCTTGTGGCATTGTGCCC 1835
Qy 1501 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1560
Db 1836 GCAAAGAGCGTGTGTGGCCCGGTATATTGCTTCACTCCCAGCCCCGTGGTGGTGGGAACG 1895
Qy 1561 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1620
Db 1896 ACCGACAGGTCGGGCGCGCCTACCTACAGCTGGGGTGCAAATGATACGGATGTCTTCGTC 1955
Qy 1621 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 1680
Db 1956 CTTAACAACACCAGGCCACCGCTGGGCAATTGGTTCGGTTGTACCTGGATGAACTCAACT 2015
Qy 1681 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 1740
Db 2016 GGATTCACCAAAGTGTGCGGAGCGCCCCCTTGTGTCATCGGAGGGGTGGGCAACAACA-- 2073
Qy 1741 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 1800
Db 2074 ----CCTTGCTCTGCCCCACTGATTGTTTCCGCAAGCATCCGGAAGCCACATACTCTCGG 2129
Qy 1801 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 1860
Db 2130 TGCGGCTCCGGTCCCTGGATTACACCCAGGTGCATGGTCGACTACCCGTATAGGCTTTGG 2189
Qy 1861 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 1920
Db 2190 CACTATCCTTGTACCATCAATTACACCATATTCAAAGTCAGGATGTACGTGGGAGGGGTC 2249
Qy 1921 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 1980
Db 2250 GAGCACAGGCTGGAAGCGGCCTGCAACTGGACGCGGGGCGAACGCTGTGATCTGGAAGAC 2309
Qy 1981 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2040
Db 2310 AGGGACAGGTCCGAGCTCAGCCCATTGCTGCTGTCCACCACACAGTGGCAGGTCCTTCCG 2369
Qy 2041 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2100
Db 2370 TGTTCTTTCACGACCCTGCCAGCCTTGTCCACCGGCCTCATCCACCTCCACCAGAACATT 2429
Qy 2101 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2160
Db 2430 GTGGACGTGCAGTACTTGTACGGGGTAGGGTCAAGCATCGCGTCCTGGGCCATTAAGTGG 2489
Qy 2161 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2220
Db 2490 GAGTACGTCGTTCTCCTGTTCCTCCTGCTTGCAGACGCGCGCGTCTGCTCCTGCTTGTGG 2549
Qy 2221 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2280
Db 2550 ATGATGTTACTCATATCCCAAGCGGAGGCGGCTTTGGAGAACCTCGTAATACTCAATGCA 2609
Qy 2281 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATATTTTGTCATCTTTTTCGTGGCTGCTTGG 2340
Db 2610 GCATCCCTGGCCGGGACGCACGGTCTTGTGTCCTTCCTCGTGTTCTTCTGCTTTGCGTGG 2669
Qy 2341 TACATCAAGGGTCGGGTAGTCCCCTTAGCTACCTATTCCCTCACTGGCCTGTGGTCCTTT 2400
Db 2670 TATCTGAAGGGTAGGTGGGTGCCCGGAGCGGTCTACGCCTTCTACGGGATGTGGCCTCTC 2729
Qy 2401 AGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTGCATGGC 2460
Db 2730 CTCCTGCTCCTGCTGGCGTTGCCTCAGCGGGCATACGCACTGGACACGGAGGTGGCCGCG 2789
Qy 2461 CAGATAGGAGCGGCTCTGCTGGTAATGATCACTCTCTTTACTCTCACCCCCGGGTATAAG 2520
Db 2790 TCGTGTGGCGGCGTTGTTCTTGTCGGGTTAATGGCGCTGACTCTGTCGCCATATTACAAG 2849
Qy 2521 ACCCTTCTCAGCCGGTTTTTGTGGTGGTTGTGCTATCTTCTGACCCTGGGGGAAGCTATG 2580
Db 2850 CGCTACATCAGCTGGTGCATGTGGTGGCTTCAGTATTTTCTGACCAGAGTAGAAGCGCAA 2909
Qy 2581 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGCATCATATGGGCC 2640
Db 2910 CTGCACGTGTGGGTTCCCCCCCTCAACGTCCGGGGGGGGCGCGATGCCGTCATCTTACTC 2969
Qy 2641 GTCGCCATATTCTACCCAGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 2700
Db 2970 ATGTGTGTTGTACACCCGACTCTGGTATTTGACATCACCAAACTACTCCTGGCCATCTTC 3029
Qy 2701 GGGCCTGCTTACCTCCTAAAAGGTGCTTTGACGCGCGTGCCGTACTTCGTCAGGGCTCAC 2760
Db 3030 GGACCCCTTTGGATTCTTCAAGCCAGTTTGCTTAAAGTCCCCTACTTCGTGCGCGTTCAA 3089
Qy 2761 GCTCTACTGAGGATGTGCACCATGGCAAGGCATCTCGCGGGGGGCAGGTACGTCCAGATG 2820
Db 3090 GGCCTTCTCCGGATCTGCGCGCTAGCGCGGAAGATAGCCGGAGGTCATTACGTGCAAATG 3149
Qy 2821 GCGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 2880
Db 3150 GCCATCATCAAGTTAGGGGCGCTTACTGGCACCTATGTGTATAACCATCTCACCCCTCTT 3209
Qy 2881 TCGGATTGGGCTGCTAGTGGCCTGCGGGACCTGGCGGTCGCCGTTGAGCCTATCATCTTC 2940
Db 3210 CGAGACTGGGCGCACAACGGCCTGCGAGATCTGGCCGTGGCTGTGGAACCAGTCGTCTTC 3269
Qy 2941 AGTCCGATGGAGAAGAAAGTCATTGTCTGGGGAGCGGAGACAGCTGCTTGTGGGGACATT 3000
Db 3270 TCCCGAATGGAGACCAAGCTCATCACGTGGGGGGCAGATACCGCCGCGTGCGGTGACATC 3329
Qy 3001 TTACACGGACTTCCCGTGTCCGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3060
Db 3330 ATCAACGGCTTGCCCGTCTCTGCCCGTAGGGGCCAGGAGATACTGCTTGGGCCAGCCGAC 3389
Qy 3061 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3120
Db 3390 GGAATGGTCTCCAAGGGGTGGAGGTTGCTGGCGCCCATCACGGCGTACGCCCAGCAGACG 3449
Qy 3121 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCT 3180
Db 3450 AGAGGCCTCCTAGGGTGTATAATCACCAGCCTGACTGGCCGGGACAAAAACCAAGTGGAG 3509
Qy 3181 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3240
Db 3510 GGTGAGGTCCAGATCGTGTCAACTGCTACCCAAACCTTCCTGGCAACGTGCATCAATGGG 3569
Qy 3241 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3300
Db 3570 GTATGCTGGACTGTCTACCACGGGGCCGGAACGAGGACCATCGCATCACCCAAGGGTCCT 3629
Qy 3301 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGG 3360
Db 3630 GTCATCCAGATGTATACCAATGTGGACCAAGACCTTGTGGGCTGGCCCGCTCCTCAAGGT 3689
Qy 3361 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3420
Db 3690 TCCCGCTCATTGACACCCTGCACCTGCGGCTCCTCGGACCTTTACCTGGTCACGAGGCAC 3749
Qy 3421 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3480
Db 3750 GCCGATGTCATTCCCGTGCGCCGGCGAGGTGATAGCAGGGGTAGCCTGCTTTCGCCCCGG 3809
Qy 3481 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3540
Db 3810 CCCATTTCCTACTTGAAAGGCTCCTCGGGGGGTCCGCTGTTGTGCCCCGCGGGACACGCC 3869
Qy 3541 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3600
Db 3870 GTGGGCCTATTCAGGGCCGCGGTGTGCACCCGTGGAGTGGCTAAGGCGGTGGACTTTATC 3929
Qy 3601 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 3660


Db 3930 CCTGTGGAGAACCTAGAGACAACCATGAGATCCCCGGTGTTCACGGACAACTCCTCTCCA 3989
Qy 3661 CCTGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 3720
Db 3990 CCAGCAGTGCCCCAGAGCTTCCAGGTGGCCCACCTGCATGCTCCCACCGGCAGCGGTAAG 4049
Qy 3721 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 3780
Db 4050 AGCACCAAGGTCCCGGCTGCGTACGCAGCCCAGGGCTACAAGGTGTTGGTGCTCAACCCC 4109
Qy 3781 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 3840
Db 4110 TCTGTTGCTGCAACGCTGGGCTTTGGTGCTTACATGTCCAAGGCCCATGGGGTTGATCCT 4169
Qy 3841 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 3900
Db 4170 AATATCAGGACCGGGGTGAGAACAATTACCACTGGCAGCCCCATCACGTACTCCACCTAC 4229
Qy 3901 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 3960
Db 4230 GGCAAGTTCCTTGCCGACGGCGGGTGCTCAGGAGGTGCTTATGACATAATAATTTGTGAC 4289
Qy 3961 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4020
Db 4290 GAGTGCCACTCCACGGATGCCACATCCATCTTGGGCATCGGCACTGTCCTTGACCAAGCA 4349
Qy 4021 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4080
Db 4350 GAGACTGCGGGGGCGAGACTGGTTGTGCTCGCCACTGCTACCCCTCCGGGCTCCGTCACT 4409
Qy 4081 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4140
Db 4410 GTGTCCCATCCTAACATCGAGGAGGTTGCTCTGTCCACCACCGGAGAGATCCCTTTTTAC 4469
Qy 4141 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4200
Db 4470 GGCAAGGCTATCCCCCTCGAGGTGATCAAGGGGGGAAGACATCTCATCTTCTGCCACTCA 4529
Qy 4201 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4260
Db 4530 AAGAAGAAGTGCGACGAGCTCGCCGCGAAGCTGGTCGCATTGGGCATCAATGCCGTGGCC 4589
Qy 4261 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4320
Db 4590 TACTACCGCGGTCTTGACGTGTCTGTCATCCCGACCAGCGGCGATGTTGTCGTCGTGTCG 4649
Qy 4321 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4380
Db 4650 ACCGATGCTCTCATGACTGGCTTTACCGGCGACTTCGACTCTGTGATAGACTGCAACACG 4709
Qy 4381 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4440
Db 4710 TGTGTCACTCAGACAGTCGATTTCAGCCTTGACCCTACCTTTACCATTGAGACAACCACG 4769
Qy 4441 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4500
Db 4770 CTCCCCCAGGATGCTGTCTCCAGGACTCAACGCCGGGGCAGGACTGGCAGGGGGAAGCCA 4829
Qy 4501 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4560
Db 4830 GGCATCTACAGATTTGTGGCACCGGGGGAGCGCCCCTCCGGCATGTTCGACTCGTCCGTC 4889
Qy 4561 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4620
Db 4890 CTCTGTGAGTGCTATGACGCGGGCTGTGCTTGGTATGAGCTCACGCCCGCCGAGACTACA 4949
Qy 4621 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 4680
Db 4950 GTTAGGCTACGAGCGTACATGAACACCCCGGGGCTTCCCGTGTGCCAGGACCATCTTGAA 5009
Qy 4681 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 4740
Db 5010 TTTTGGGAGGGCGTCTTTACGGGCCTCACTCATATAGATGCCCACTTTCTATCCCAGACA 5069
Qy 4741 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 4800
Db 5070 AAGCAGAGTGGGGAGAACTTTCCTTACCTGGTAGCGTACCAAGCCACCGTGTGCGCTAGG 5129
Qy 4801 GCCAAAGCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 4860
Db 5130 GCTCAAGCCCCTCCCCCATCGTGGGACCAGATGTGGAAGTGTTTGATCCGCCTTAAACCC 5189
Qy 4861 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 4920
Db 5190 ACCCTCCATGGGCCAACACCCCTGCTATACAGACTGGGCGCTGTTCAGAATGAAGTCACC 5249
Qy 4921 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 4980
Db 5250 CTGACGCACCCAATCACCAAATACATCATGACATGCATGTCGGCCGACCTGGAGGTCGTC 5309
Qy 4981 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5040
Db 5310 ACGAGCACCTGGGTGCTCGTTGGCGGCGTCCTGGCTGCTCTGGCCGCGTATTGCCTGTCA 5369
Qy 5041 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5100
Db 5370 ACAGGCTGCGTGGTCATAGTGGGCAGGATTGTCTTGTCCGGGAAGCCGGCAATTATACCT 5429
Qy 5101 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5160
Db 5430 GACAGGGAGGTTCTCTACCAGGAGTTCGATGAGATGGAAGAGTGCTCTCAGCACTTACCG 5489
Qy 5161 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5220
Db 5490 TACATCGAGCAAGGGATGATGCTCGCTGAGCAGTTCAAGCAGAAGGCCCTCGGCCTCCTG 5549
Qy 5221 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5280
Db 5550 CAGACCGCGTCCCGCCAAGCAGAGGTTATCACCCCTGCTGTCCAGACCAACTGGCAGAAA 5609
Qy 5281 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5340
Db 5610 CTCGAGGTCTTCTGGGCGAAGCACATGTGGAATTTCATCAGTGGGATACAATACTTGGCG 5669
Qy 5341 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5400
Db 5670 GGCCTGTCAACGCTGCCTGGTAACCCCGCCATTGCTTCATTGATGGCTTTTACAGCTGCC 5729
Qy 5401 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5460
Db 5730 GTCACCAGCCCACTAACCACTGGCCAAACCCTCCTCTTCAACATATTGGGGGGGTGGGTG 5789
Qy 5461 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5520
Db 5790 GCTGCCCAGCTCGCCGCCCCCGGTGCCGCTACCGCCTTTGTGGGCGCTGGCTTAGCTGGC 5849
Qy 5521 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5580
Db 5850 GCCGCCATCGGCAGCGTTGGACTGGGGAAGGTCCTCGTGGACATTCTTGCAGGGTATGGC 5909
Qy 5581 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5640
Db 5910 GCGGGCGTGGCGGGAGCTCTTGTAGCATTCAAGATCATGAGCGGTGAGGTCCCCTCCACG 5969
Qy 5641 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 5700
Db 5970 GAGGACCTGGTCAATCTGCTGCCCGCCATCCTCTCGCCTGGAGCCCTTGTAGTCGGTGTG 6029
Qy 5701 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 5760
Db 6030 GTCTGCGCAGCAATACTGCGCCGGCACGTTGGCCCGGGCGAGGGGGCAGTGCAATGGATG 6089
Qy 5761 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 5820
Db 6090 AACCGGCTAATAGCCTTCGCCTCCCGGGGGAACCATGTTTCCCCCACGCACTACGTGCCG 6149



```
Db       8244 ATGGGGTTCTCGTATGATACCCGCTGTTTTGACTCCACAGTCACTGAGAGCGACATCCGT 8303

Qy       8029 ACTGAGGAGTCCATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACAC 8088
             || |||||| | || || |      ||   | ||  |    | |||| |   ||||| | |
Db       8304 ACGGAGGAGGCAATTTACCAATGTTGTGACCTGGACCCCCAAGCCCGCGTGGCCATCAAG 8363

Qy       8089 TCGCTAACTGAGAGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGC 8148
             || || |||||||| ||||| || || || ||| |   |||    | |||  ||| ||||
Db       8364 TCCCTCACTGAGAGGCTTTATGTTGGGGGCCCTCTTACCAATTCAAGGGGGGAAAACTGC 8423

Qy       8149 GGGTACAGGCGTTGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACA 8208
             || ||| |   | |||||||| ||||| || || || ||||||   || |||||| ||||
Db       8424 GGCTACCGCAGGTGCCGCGCGAGCGGCGTACTGACAACTAGCTGTGGTAACACCCTCACT 8483

Qy       8209 TGCTACGTGAAAGCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTG 8268
             |||||| | || |||    || || |||  ||| |||||| |     |     ||| |||||
Db       8484 TGCTACATCAAGGCCCGGGCAGCCTGTCGAGCCGCAGGGCTCCAGGACTGCACCATGCTC 8543

Qy       8269 GTATGCGGCGATGACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGG 8328
             || || ||||| ||||| || || |||| |||||   ||||  | ||||||||| |  |
Db       8544 GTGTGTGGCGACGACTTAGTCGTTATCTGTGAAAGTGCGGGGGTCCAGGAGGACGCGGCG 8603

Qy       8329 AACCTGAGAGCCTTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCC 8388
             | |||||||||||||||||||||||||||||||| || ||||| || || ||||||||
Db       8604 AGCCTGAGAGCCTTCACGGAGGCTATGACCAGGTACTCCGCCCCCCCCGGGGACCCCCCA 8663

Qy       8389 AGACCGGAGTATGATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTG 8448
                ||| || || ||  ||||||| |||||||| |||||||| || ||||| || || |
Db       8664 CAACCAGAATACGACTTGGAGCTTATAACATCATGCTCCTCCAACGTGTCAGTCGCCCAC 8723

Qy       8449 GGCCCACAAGGCCGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGG 8508
             | |      ||      |  |||||||| ||| | |||||||| || ||   ||||  |
Db       8724 GACGGCGCTGGAAAGAGGGTCTACTACCTTACCCGTGACCCTACAACCCCCCTCGCGAGA 8783

Qy       8509 GCTGCCTGGGAAACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAG 8568
             || || ||||| ||||  |||||| | || |||||||| ||||| || ||||| |||  |
Db       8784 GCCGCGTGGGAGACAGCAAGACACACTCCAGTCAATTCCTGGCTAGGCAACATAATCATG 8843

Qy       8569 TACGCCCCGACCATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATG 8628
             |  ||||| ||  | ||||  | ||| | |||||||| || |||||    | | |||||
Db       8844 TTTGCCCCCACACTGTGGGCGAGGATGATACTGATGACCCATTTCTTTAGCGTCCTCATA 8903

Qy       8629 GCTCAAGACACGCTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTG 8688
             ||    ||   ||| || |||    || |||| |||||||    |||||| |
Db       8904 GCCAGGGATCAGCTTGAACAGGCTCTTAACTGTGAGATCTACGGAGCCTGCTACTCCATA 8963

Qy       8689 AGTCCCTTGGACCTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTG 8748
                ||  |||| || ||  | || ||| ||||  | ||||| ||   ||| ||||| ||
Db       8964 GAACCACTGGATCTACCTCCAATCATTCAAAGACTCCATGGCCTCAGCGCATTTTCACTC 9023

Qy       8749 CACACATACACTCCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCG 8808
             ||||  ||| ||||    ||| | |   |||||||  ||  |||||||||||||||||
Db       9024 CACAGTTACTCTCCAGGTGAAATCAATAGGGTGGCCGCATGCCTCAGAAAACTTGGGGTC 9083

Qy       8809 CCACCCCTCAGAGCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGT 8868
             || ||| |  |||| ||||     |||||| ||    ||| | ||   ||  | ||| |
Db       9084 CCGCCCTTGCGAGCTTGGAGACACCGGGCCCGGAGCGTCCGCGCTAGGCTTCTGTCCAGA 9143

Qy       8869 GGGGGGAGAGCGGCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTC 8928
             || || || || ||| | || ||     || |||||||| |||||  || |  || ||||||
Db       9144 GGAGGCAGGGCTGCCATATGTGGCAAGTACCTCTTCAACTGGGCAGTAAGAACAAAGCTC 9203

Qy       8929 AAACTCACTCCATTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGC 8988
             |||||||||||  |  |||  ||  |||  ||||| || ||| | |||||||| |  |||
Db       9204 AAACTCACTCCAATAGCGGCCGCTGGCCGGCTGGACTTGTCCGGTTGGTTCACGGCTGGC 9263

Qy       8989 GCCGGCGGGGGCGACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTT 9048
                | |||||||  |||||||||||||||||||| | |||||| |||||||    ||    |||
Db       9264 TACAGCGGGGGAGACATTTATCACAGCGTGTCTCATGCCCGGCCCCGCTGGTTCTGGTTT 9323

Qy       9049 GGCCTACTCCTACTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCG 9098
              |||||||||| ||   || |||||||||| | | ||| ||||||   ||
Db       9324 TGCCTACTCCTGCTCGCTGCAGGGGTAGGCATCTACCTCCTCCCCAACCG 9373
```

```
RESULT 11
US-09-014-416-2
; Sequence 2, Application US/09014416
; Patent No. 6153421
; GENERAL INFORMATION:
;  APPLICANT: Yanagi, Masayuki
;  APPLICANT: Bukh, Jens
;  APPLICANT: Emerson, Susanne U.
;  APPLICANT: Purcell, Robert H.
;  TITLE OF INVENTION: CLONED GENOMES OF INFECTIOUS HEPATITIS C VIRUSES AND
;  TITLE OF INVENTION: USES THEREOF
;  FILE REFERENCE: 20264276
;  CURRENT APPLICATION NUMBER: US/09/014,416
;  CURRENT FILING DATE: 1998-01-27
;  EARLIER APPLICATION NUMBER: US 60/053,062
;  EARLIER FILING DATE: 1997-07-18
;  NUMBER OF SEQ ID NOS: 65
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 9599
;   TYPE: DNA
;   ORGANISM: Hepatitis C virus
US-09-014-416-2

  Query Match            45.3%;  Score 4121.6;  DB 3;  Length 9599;
  Best Local Similarity  66.4%;  Pred. No. 0;
  Matches 6046;  Conservative    0;  Mismatches 2974;  Indels   90;  Gaps    5;

Qy          1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 60
              |||||||| |||||||||||||||||||||||||| | ||||||||||||||||||||
Db        342 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGTCGCCCACAG 401

Qy         61 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 120
              ||||| ||||| ||||| ||||| |||||||||| |||||| ||||||||||||||||||
Db        402 GACGTCAAGTTCCCGGGTGGCGGTCAGATCGTTGGTGGAGTTTACTTGTTGCCGCGCAGG 461

Qy        121 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 180
              ||||| || |||||||||||||||| |||||||||| |||||||| || || || ||
Db        462 GGCCCTAGATTGGGTGTGCGCGCGACGAGGAAGACTTCCGAGCGGTCGCAACCTCGAGGT 521

Qy        181 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 240
              || || ||||| ||||| || |  || ||  ||  |||| ||||||   | || ||
Db        522 AGACGTCAGCCTATCCCCAAGGCACGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGG 581

Qy        241 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 300
              ||||| |||||||| || || ||||||||    ||| ||||| |||||||||||||| |||
Db        582 TACCCTTGGCCCCTCTATGGCAATGAGGGTTGCGGGTGGGCGGGATGGCTCCTGTCTCCC 641

Qy        301 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 360
              || || || || ||    ||||||||||  ||||||||| ||||||||||   |||||
Db        642 CGTGGCTCTCGGCCTAGCTGGGGCCCCACAGACCCCCGGCGTAGGTCGCGCAATTTGGGT 701

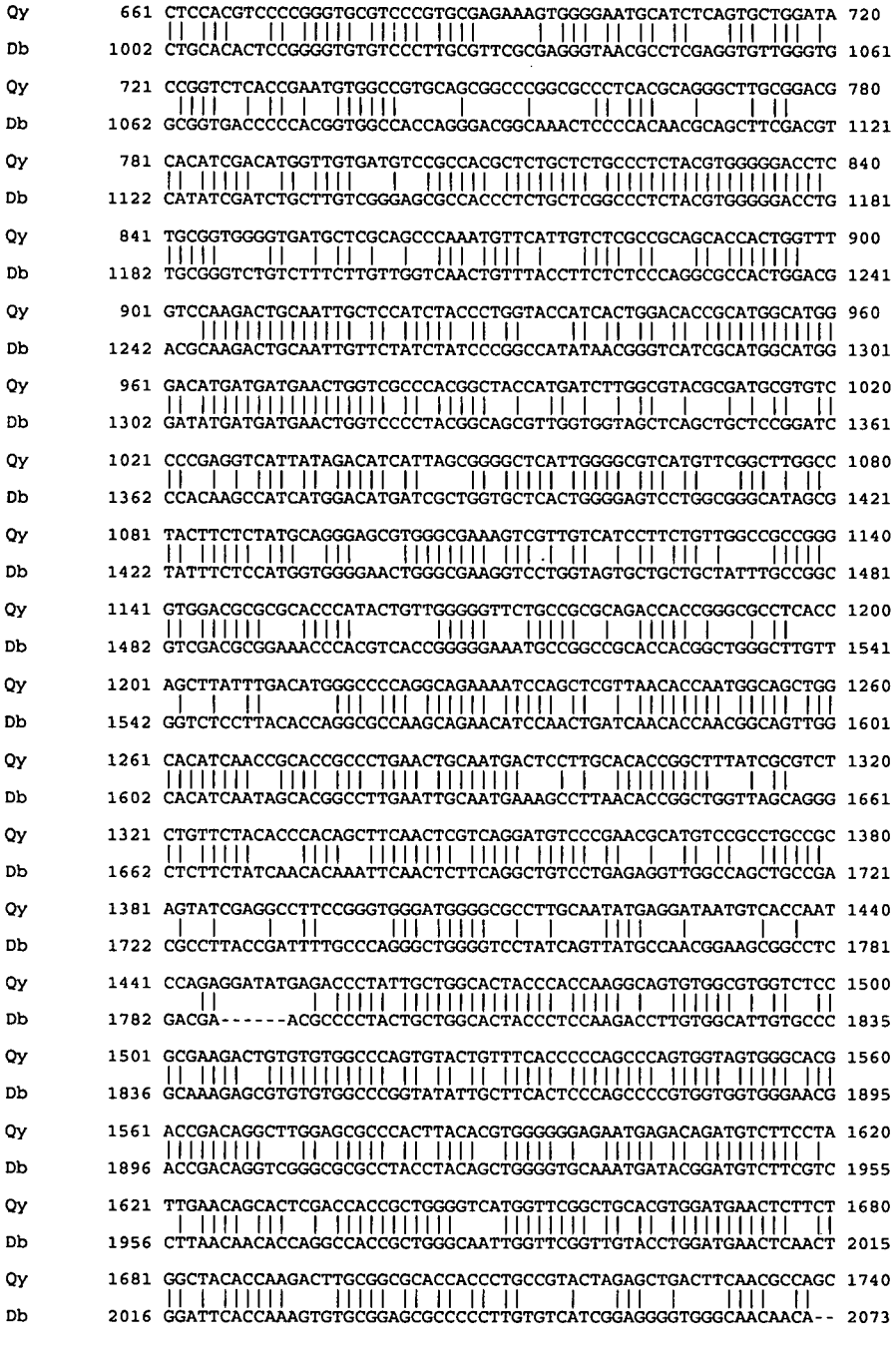
Qy        361 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 420
              |||||||||||||||| |||||||||| |||||||||||||||||| ||   ||||
Db        702 AAGGTCATCGATACCCTTACGTGCGGCTTCGCCGACCTCATGGGGTACATACCGCTCGTC 761

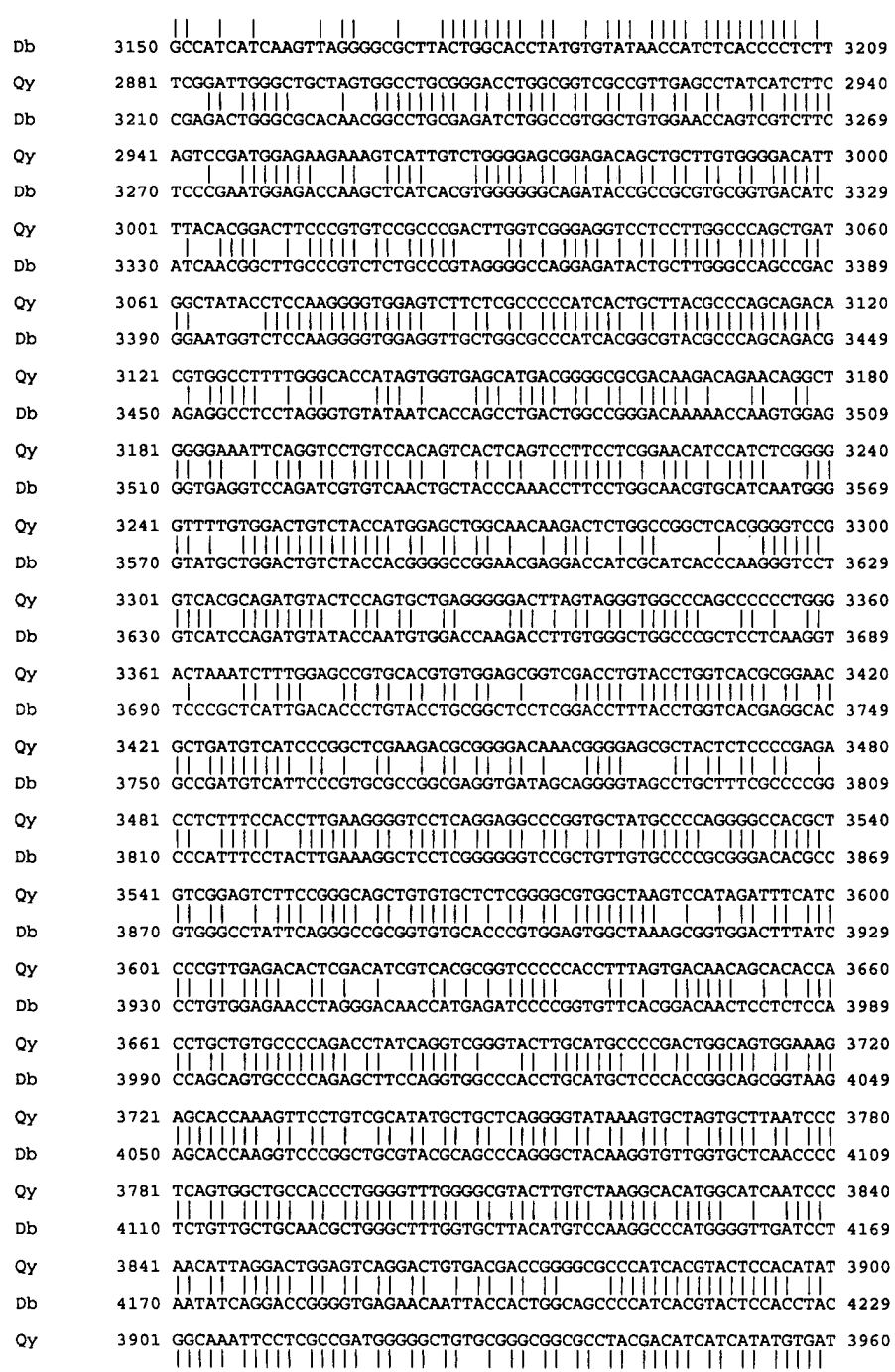
Qy        421 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 480
              |||||||| || || ||||  ||||| || || |||||||||||  | || ||||| |||
Db        762 GGCGCCCCTCTTGGAGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAAGAC 821

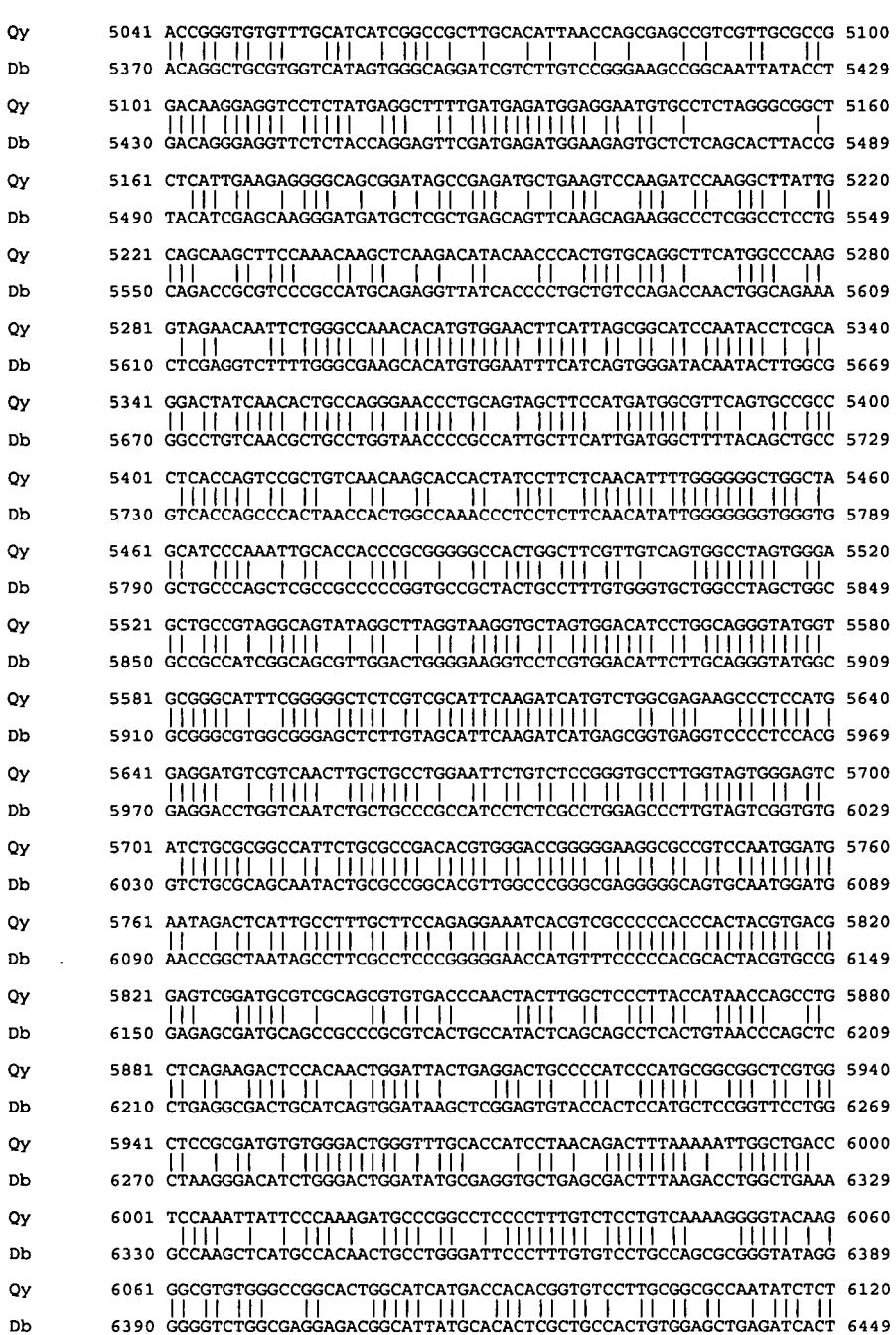
Qy        481 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 540
              || || || | |||||||||||| | || |||||||| || |||||||| | ||||||
Db        822 GGCGTGAACTATGCAACAGGGAACCTTCCTGGTTGCTCTTTCTCTATCTTCCTTCTGGCC 881

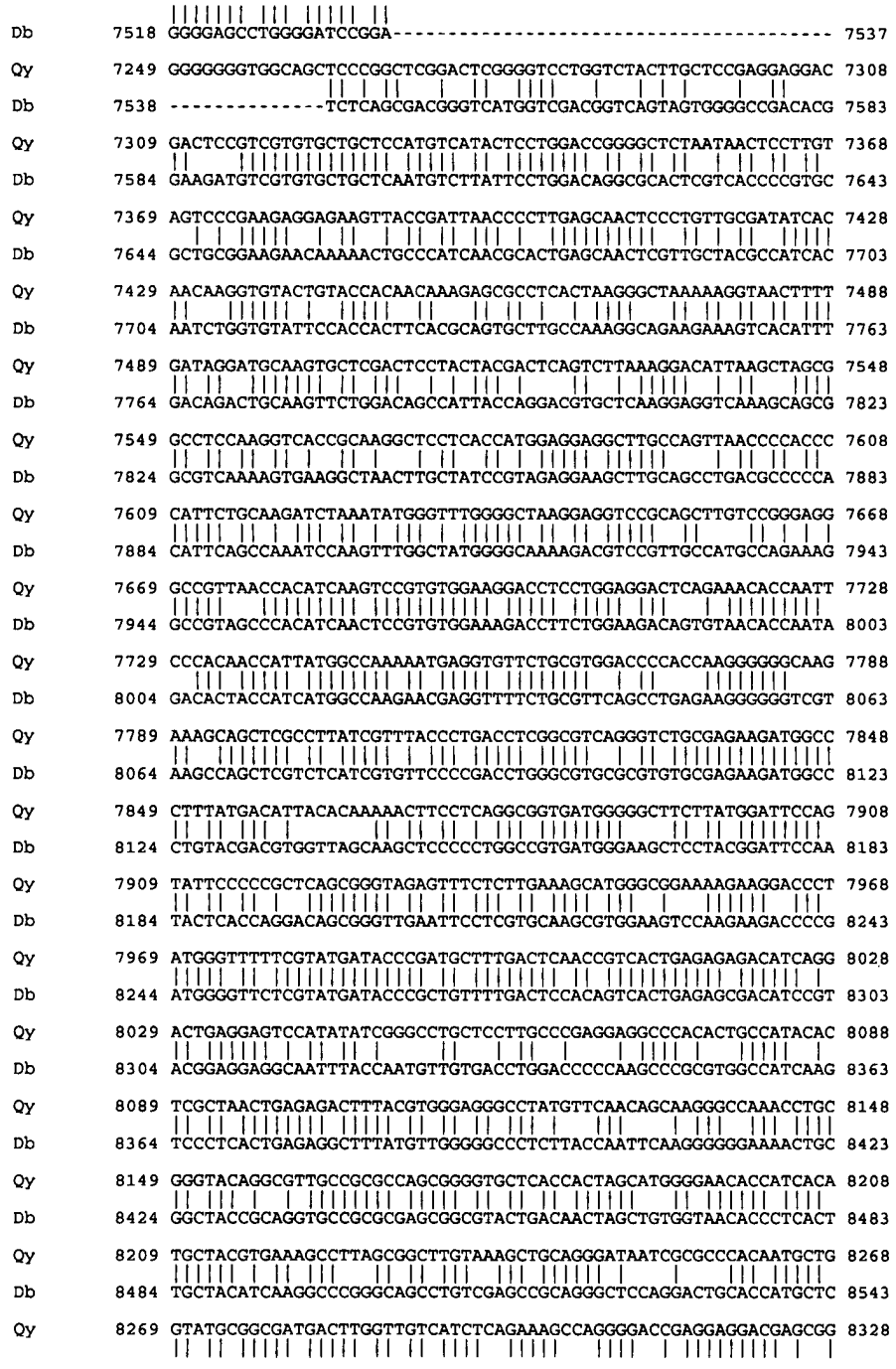
Qy        541 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 600
              ||||| || ||| | ||    || |  || ||    | |||||    ||  |
Db        882 CTGCTCTCTTGCCTGACTGTGCCCGCTTCAGCCTACCAAGTGCGCAATTCCTCGGGGCTT 941

Qy        601 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 660
              |||   || || || || ||| | ||     || |||    |   ||       | |||  ||
Db        942 TACCATGTCACCAATGATTGCCCTAACTCGAGTATTGTGTACGAGGCGGCCGATGCCATC 1001
```


Qy 661 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 720
Db 1002 CTGCACACTCCGGGGTGTGTCCCTTGCGTTCGCGAGGGTAACGCCTCGAGGTGTTGGGTG 1061
Qy 721 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 780
Db 1062 GCGGTGACCCCCACGGTGGCCACCAGGGACGGCAAACTCCCCACAACGCAGCTTCGACGT 1121
Qy 781 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 840
Db 1122 CATATCGATCTGCTTGTCGGGAGCGCCACCCTCTGCTCGGCCCTCTACGTGGGGGACCTG 1181
Qy 841 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 900
Db 1182 TGCGGGTCTGTCTTTCTTGTTGGTCAACTGTTTACCTTCTCTCCCAGGCGCCACTGGACG 1241
Qy 901 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 960
Db 1242 ACGCAAGACTGCAATTGTTCTATCTATCCCGGCCATATAACGGGTCATCGCATGGCATGG 1301
Qy 961 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1020
Db 1302 GATATGATGATGAACTGGTCCCCTACGGCAGCGTTGGTGGTAGCTCAGCTGCTCCGGATC 1361
Qy 1021 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1080
Db 1362 CCACAAGCCATCATGGACATGATCGCTGGTGCTCACTGGGGAGTCCTGGCGGGCATAGCG 1421
Qy 1081 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1140
Db 1422 TATTTCTCCATGGTGGGGAACTGGGCGAAGGTCCTGGTAGTGCTGCTGCTATTTGCCGGC 1481
Qy 1141 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1200
Db 1482 GTCGACGCGGAAACCCACGTCACCGGGGGAAATGCCGGCCGCACCACGGCTGGGCTTGTT 1541
Qy 1201 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1260
Db 1542 GGTCTCCTTACACCAGGCGCCAAGCAGAACATCCAACTGATCAACACCAACGGCAGTTGG 1601
Qy 1261 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1320
Db 1602 CACATCAATAGCACGGCCTTGAATTGCAATGAAAGCCTTAACACCGGCTGGTTAGCAGGG 1661
Qy 1321 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1380
Db 1662 CTCTTCTATCAACACAAATTCAACTCTTCAGGCTGTCCTGAGAGGTTGGCCAGCTGCCGA 1721
Qy 1381 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1440
Db 1722 CGCCTTACCGATTTTGCCCAGGGCTGGGGTCCTATCAGTTATGCCAACGGAAGCGGCCTC 1781
Qy 1441 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1500
Db 1782 GACGA------ACGCCCCTACTGCTGGCACTACCCTCCAAGACCTTGTGGCATTGTGCCC 1835
Qy 1501 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1560
Db 1836 GCAAAGAGCGTGTGTGGCCCGGTATATTGCTTCACTCCCAGCCCCGTGGTGGTGGGAACG 1895
Qy 1561 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1620
Db 1896 ACCGACAGGTCGGGCGCGCCTACCTACAGCTGGGGTGCAAATGATACGGATGTCTTCGTC 1955
Qy 1621 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 1680
Db 1956 CTTAACAACACCAGGCCACCGCTGGGCAATTGGTTCGGTTGTACCTGGATGAACTCAACT 2015
Qy 1681 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 1740
Db 2016 GGATTCACCAAAGTGTGCGGAGCGCCCCCTTGTGTCATCGGAGGGGTGGGCAACAACA-- 2073
Qy 1741 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 1800
Db 2074 ----CCTTGCTCTGCCCCACTGATTGCTTCCGCAAACATCCGGAAGCCACATACTCTCGG 2129
Qy 1801 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 1860
Db 2130 TGCGGCTCCGGTCCCTGGATTACACCCAGGTGCATGGTCGACTACCCGTATAGGCTTTGG 2189
Qy 1861 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 1920
Db 2190 CACTATCCTTGTACCATCAATTACACCATATTCAAAGTCAGGATGTACGTGGGAGGGGTC 2249
Qy 1921 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 1980
Db 2250 GAGCACAGGCTGGAAGCGGCCTGCAACTGGACGCGGGGCGAACGCTGTGATCTGGAAGAC 2309
Qy 1981 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2040
Db 2310 AGGGACAGGTCCGAGCTCAGCCCGTTGCTGCTGTCCACCACACAGTGGCAGGTCCTTCCG 2369
Qy 2041 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2100
Db 2370 TGTTCTTTCACGACCCTGCCAGCCTTGTCCACCGGCCTCATCCACCTCCACCAGAACATT 2429
Qy 2101 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2160
Db 2430 GTGGACGTGCAGTACTTGTACGGGGTAGGGTCAAGCATCGCGTCCTGGGCCATTAAGTGG 2489
Qy 2161 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2220
Db 2490 GAGTACGTCGTTCTCCTGTTCCTTCTGCTTGCAGACGCGCGCGTCTGCTCCTGCTTGTGG 2549
Qy 2221 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2280
Db 2550 ATGATGTTACTCATATCCCAAGCGGAGGCGGCTTTGGAGAACCTCGTAATACTCAATGCA 2609
Qy 2281 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATATTTTGTCATCTTTTTCGTGGCTGCTTGG 2340
Db 2610 GCATCCCTGGCCGGGACGCACGGTCTTGTGTCCTTCCTCGTGTTCTTCTGCTTTGCGTGG 2669
Qy 2341 TACATCAAGGGTCGGGTAGTCCCCTTAGCTACCTATTCCCTCACTGGCCTGTGGTCCTTT 2400
Db 2670 TATCTGAAGGGTAGGTGGGTGCCCGGAGCGGTCTACGCCCTCTACGGGATGTGGCCTCTC 2729
Qy 2401 AGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTGCATGGC 2460
Db 2730 CTCCTGCTCCTGCTGGCGTTGCCTCAGCGGGCATACGCACTGGACACGGAGGTGGCCGCG 2789
Qy 2461 CAGATAGGAGCGGCTCTGCTGGTAATGATCACTCTCTTTACTCTCACCCCCGGGTATAAG 2520
Db 2790 TCGTGTGGCGGCGTTGTTCTTGTCGGGTTAATGGCGCTGACTCTGTCGCCATATTACAAG 2849
Qy 2521 ACCCTTCTCAGCCGGTTTTTGTGGTGGTTGTGCTATCTTCTGACCCTGGGGGAAGCTATG 2580
Db 2850 CGCTATATCAGCTGGTGCATGTGGTGGCTTCAGTATTTTCTGACCAGAGTAGAAGCGCAA 2909
Qy 2581 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGCATCATATGGGCC 2640
Db 2910 CTGCACGTGTGGGTTCCCCCCCTCAACGTCCGGGGGGGGCGCGATGCCGTCATCTTACTC 2969
Qy 2641 GTCGCCATATTCTACCCAGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 2700
Db 2970 ATGTGTGTAGTACACCCGACCCTGGTATTTGACATCACCAAACTACTCCTGGCCATCTTC 3029
Qy 2701 GGGCCTGCTTACCTCCTAAAAGGTGCTTTGACGCGCGTGCCGTACTTCGTCAGGGCTCAC 2760
Db 3030 GGACCCCTTTGGATTCTTCAAGCCAGTTTGCTTAAAGTCCCCTACTTCGTGCGCGTTCAA 3089
Qy 2761 GCTCTACTGAGGATGTGCACCATGGCAAGGCATCTCGCGGGGGGCAGGTACGTCCAGATG 2820
Db 3090 GGCCTTCTCCGGATCTGCGCGCTAGCGCGGAAGATAGCCGGAGGTCATTACGTGCAAATG 3149
Qy 2821 GCGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 2880


Db 3150 GCCATCATCAAGTTAGGGGCGCTTACTGGCACCTATGTGTATAACCATCTCACCCCTCTT 3209
Qy 2881 TCGGATTGGGCTGCTAGTGGCCTGCGGGACCTGGCGGTCGCCGTTGAGCCTATCATCTTC 2940
Db 3210 CGAGACTGGGCGCACAACGGCCTGCGAGATCTGGCCGTGGCTGTGGAACCAGTCGTCTTC 3269
Qy 2941 AGTCCGATGGAGAAGAAAGTCATTGTCTGGGGAGCGGAGACAGCTGCTTGTGGGGACATT 3000
Db 3270 TCCCGAATGGAGACCAAGCTCATCACGTGGGGGGCAGATACCGCCGCGTGCGGTGACATC 3329
Qy 3001 TTACACGGACTTCCCGTGTCCGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3060
Db 3330 ATCAACGGCTTGCCCGTCTCTGCCCGTAGGGGCCAGGAGATACTGCTTGGGCCAGCCGAC 3389
Qy 3061 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3120
Db 3390 GGAATGGTCTCCAAGGGGTGGAGGTTGCTGGCGCCCATCACGGCGTACGCCCAGCAGACG 3449
Qy 3121 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCT 3180
Db 3450 AGAGGCCTCCTAGGGTGTATAATCACCAGCCTGACTGGCCGGGACAAAAACCAAGTGGAG 3509
Qy 3181 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3240
Db 3510 GGTGAGGTCCAGATCGTGTCAACTGCTACCCAAACCTTCCTGGCAACGTGCATCAATGGG 3569
Qy 3241 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3300
Db 3570 GTATGCTGGACTGTCTACCACGGGGCCGGAACGAGGACCATCGCATCACCCAAGGGTCCT 3629
Qy 3301 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCTGGG 3360
Db 3630 GTCATCCAGATGTATACCAATGTGGACCAAGACCTTGTGGGCTGGCCCGCTCCTCAAGGT 3689
Qy 3361 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3420
Db 3690 TCCCGCTCATTGACACCCTGTACCTGCGGCTCCTCGGACCTTTACCTGGTCACGAGGCAC 3749
Qy 3421 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3480
Db 3750 GCCGATGTCATTCCCGTGCGCCGGCGAGGTGATAGCAGGGGTAGCCTGCTTTCGCCCCGG 3809
Qy 3481 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3540
Db 3810 CCCATTTCCTACTTGAAAGGCTCCTCGGGGGGTCCGCTGTTGTGCCCCGCGGGACACGCC 3869
Qy 3541 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3600
Db 3870 GTGGGCCTATTCAGGGCCGCGGTGTGCACCCGTGGAGTGGCTAAAGCGGTGGACTTTATC 3929
Qy 3601 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 3660
Db 3930 CCTGTGGAGAACCTAGGGACAACCATGAGATCCCCGGTGTTCACGGACAACTCCTCTCCA 3989
Qy 3661 CCTGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 3720
Db 3990 CCAGCAGTGCCCCAGAGCTTCCAGGTGGCCCACCTGCATGCTCCCACCGGCAGCGGTAAG 4049
Qy 3721 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 3780
Db 4050 AGCACCAAGGTCCCGGCTGCGTACGCAGCCCAGGGCTACAAGGTGTTGGTGCTCAACCCC 4109
Qy 3781 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 3840
Db 4110 TCTGTTGCTGCAACGCTGGGCTTTGGTGCTTACATGTCCAAGGCCCATGGGGTTGATCCT 4169
Qy 3841 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 3900
Db 4170 AATATCAGGACCGGGGTGAGAACAATTACCACTGGCAGCCCCATCACGTACTCCACCTAC 4229
Qy 3901 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 3960
Db 4230 GGCAAGTTCCTTGCCGACGGCGGGTGCTCAGGAGGTGCTTATGACATAATAATTTGTGAC 4289
Qy 3961 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4020
Db 4290 GAGTGCCACTCCACGGATGCCACATCCATCTTGGGCATCGGCACTGTCCTTGACCAAGCA 4349
Qy 4021 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4080
Db 4350 GAGACTGCGGGGGCGAGACTGGTTGTGCTCGCCACTGCTACCCCTCCGGGCTCCGTCACT 4409
Qy 4081 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4140
Db 4410 GTGTCCCATCCTAACATCGAGGAGGTTGCTCTGTCCACCACCGGAGAGATCCCCTTTTAC 4469
Qy 4141 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4200
Db 4470 GGCAAGGCTATCCCCCTCGAGGTGATCAAGGGGGGAAGACATCTCATCTTCTGCCACTCA 4529
Qy 4201 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4260
Db 4530 AAGAAGAAGTGCGACGAGCTCGCCGCGAAGCTGGTCGCATTGGGCATCAATGCCGTGGCC 4589
Qy 4261 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4320
Db 4590 TACTACCGCGGTCTTGACGTGTCTGTCATCCCGACCAGCGGCGATGTTGTCGTCGTGTCG 4649
Qy 4321 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4380
Db 4650 ACCGATGCTCTCATGACTGGCTTTACCGGCGACTTCGACTCTGTGATAGACTGCAACACG 4709
Qy 4381 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4440
Db 4710 TGTGTCACTCAGACAGTCGATTTCAGCCTTGACCCTACCTTTACCATTGAGACAACCACG 4769
Qy 4441 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4500
Db 4770 CTCCCCCAGGATGCTGTCTCCAGGACTCAACGCCGGGGCAGGACTGGCAGGGGGAAGCCA 4829
Qy 4501 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4560
Db 4830 GGCATCTATAGATTTGTGGCACCGGGGGAGCGCCCCTCCGGCATGTTCGACTCGTCCGTC 4889
Qy 4561 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4620
Db 4890 CTCTGTGAGTGCTATGACGCGGGCTGTGCTTGGTATGAGCTCACGCCCGCCGAGACTACA 4949
Qy 4621 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 4680
Db 4950 GTTAGGCTACGAGCGTACATGAACACCCCGGGGCTTCCCGTGTGCCAGGACCATCTTGAA 5009
Qy 4681 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 4740
Db 5010 TTTTGGGAGGGCGTCTTTACGGGCCTCACTCATATAGATGCCCACTTTTTATCCCAGACA 5069
Qy 4741 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 4800
Db 5070 AAGCAGAGTGGGGAGAACTTTCCTTACCTGGTAGCGTACCAAGCCACCGTGTGCGCTAGG 5129
Qy 4801 GCCAAAGCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 4860
Db 5130 GCTCAAGCCCCTCCCCCATCGTGGGACCAGATGTGGAAGTGTTTGATCCGCCTTAAACCC 5189
Qy 4861 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 4920
Db 5190 ACCCTCCATGGGCCAACACCCCTGCTATACAGACTGGGCGCTGTTCAGAATGAAGTCACC 5249
Qy 4921 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 4980
Db 5250 CTGACGCACCCAATCACCAAATACATCATGACATGCATGTCGGCCGACCTGGAGGTCGTC 5309
Qy 4981 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5040
Db 5310 ACGAGCACCTGGGTGCTCGTTGGCGGCGTCCTGGCTGCTCTGGCCGCGTATTGCCTGTCA 5369


Qy 5041 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5100
Db 5370 ACAGGCTGCGTGGTCATAGTGGGCAGGATCGTCTTGTCCGGGAAGCCGGCAATTATACCT 5429
Qy 5101 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5160
Db 5430 GACAGGGAGGTTCTCTACCAGGAGTTCGATGAGATGGAAGAGTGCTCTCAGCACTTACCG 5489
Qy 5161 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5220
Db 5490 TACATCGAGCAAGGGATGATGCTCGCTGAGCAGTTCAAGCAGAAGGCCCTCGGCCTCCTG 5549
Qy 5221 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5280
Db 5550 CAGACCGCGTCCCGCCATGCAGAGGTTATCACCCCTGCTGTCCAGACCAACTGGCAGAAA 5609
Qy 5281 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5340
Db 5610 CTCGAGGTCTTTTGGGCGAAGCACATGTGGAATTTCATCAGTGGGATACAATACTTGGCG 5669
Qy 5341 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5400
Db 5670 GGCCTGTCAACGCTGCCTGGTAACCCCGCCATTGCTTCATTGATGGCTTTTACAGCTGCC 5729
Qy 5401 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5460
Db 5730 GTCACCAGCCCACTAACCACTGGCCAAACCCTCCTCTTCAACATATTGGGGGGGTGGGTG 5789
Qy 5461 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5520
Db 5790 GCTGCCCAGCTCGCCGCCCCCGGTGCCGCTACTGCCTTTGTGGGTGCTGGCCTAGCTGGC 5849
Qy 5521 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5580
Db 5850 GCCGCCATCGGCAGCGTTGGACTGGGGAAGGTCCTCGTGGACATTCTTGCAGGGTATGGC 5909
Qy 5581 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5640
Db 5910 GCGGGCGTGGCGGGAGCTCTTGTAGCATTCAAGATCATGAGCGGTGAGGTCCCCTCCACG 5969
Qy 5641 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 5700
Db 5970 GAGGACCTGGTCAATCTGCTGCCCGCCATCCTCTCGCCTGGAGCCCTTGTAGTCGGTGTG 6029
Qy 5701 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 5760
Db 6030 GTCTGCGCAGCAATACTGCGCCGGCACGTTGGCCCGGGCGAGGGGGCAGTGCAATGGATG 6089
Qy 5761 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 5820
Db 6090 AACCGGCTAATAGCCTTCGCCTCCCGGGGGAACCATGTTTCCCCCACGCACTACGTGCCG 6149
Qy 5821 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 5880
Db 6150 GAGAGCGATGCAGCCGCCCGCGTCACTGCCATACTCAGCAGCCTCACTGTAACCCAGCTC 6209
Qy 5881 CTCAGAAGACTCCACAACTGGATTACTGAGGACTGCCCCATCCCATGCGGCGGCTCGTGG 5940
Db 6210 CTGAGGCGACTGCATCAGTGGATAAGCTCGGAGTGTACCACTCCATGCTCCGGTTCCTGG 6269
Qy 5941 CTCCGCGATGTGTGGGACTGGGTTTGCACCATCCTAACAGACTTTAAAAATTGGCTGACC 6000
Db 6270 CTAAGGGACATCTGGGACTGGATATGCGAGGTGCTGAGCGACTTTAAGACCTGGCTGAAA 6329
Qy 6001 TCCAAATTATTCCCAAAGATGCCCGGCCTCCCCTTTGTCTCCTGTCAAAAGGGGTACAAG 6060
Db 6330 GCCAAGCTCATGCCACAACTGCCTGGGATTCCCTTTGTGTCCTGCCAGCGCGGGTATAGG 6389
Qy 6061 GGCGTGTGGGCCGGCACTGGCATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6120
Db 6390 GGGGTCTGGCGAGGAGACGGCATTATGCACACTCGCTGCCACTGTGGAGCTGAGATCACT 6449
Qy 6121 GGCAATGTCCGCTTGGGCTCCATGAGAATCACGGGGCCTAAGACCTGCATGAATATCTGG 6180
Db 6450 GGACATGTCAAAAACGGGACGATGAGGATCGTCGGTCCTAGGACCTGCAGGAACATGTGG 6509
Qy 6181 CAGGGGACCTTTCCTATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCGCCA 6240
Db 6510 AGTGGGACGTTCCCCATTAACGCCTACACCACGGGCCCCTGTACTCCCCTTCCTGCGCCG 6569
Qy 6241 AACTTTAAGGTCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6300
Db 6570 AACTATAAGTTCGCGCTGTGGAGGGTGTCTGCAGAGGAATACGTGGAGATAAGGCGGGTG 6629
Qy 6301 GGGTCATACCACTACATAACAGGACTCACCACTGATAACTTGAAAGTCCCCTGCCAACTA 6360
Db 6630 GGGGACTTCCACTACGTATCGGGTATGACTACTGACAATCTTAAATGCCCGTGCCAGATC 6689
Qy 6361 CCCTCTCCCGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCACA 6420
Db 6690 CCATCGCCCGAATTTTTCACAGAATTGGACGGGGTGCGCCTACACAGGTTTGCGCCCCCT 6749
Qy 6421 CCGAAGCCGTTTTTCCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6480
Db 6750 TGCAAGCCCTTGCTGCGGGAGGAGGTATCATTCAGAGTAGGACTCCACGAGTACCCGGTG 6809
Qy 6481 GGGTCCCAGCTTCCTTGCGACCCTGAACCCGACACAGACGTATTGATGTCCATGCTAACA 6540
Db 6810 GGGTCGCAATTACCTTGCGAGCCCGAACCGGACGTAGCCGTGTTGACGTCCATGCTCACT 6869
Qy 6541 GATCCATCTCATATCACGGCGGAGACTGCAGCGCGGCGTTTAGCGCGGGGGTCACCCCCA 6600
Db 6870 GATCCCTCCCATATAACAGCAGAGGCGGCCGGGAGAAGGTTGGCGAGAGGGTCACCCCCT 6929
Qy 6601 TCCGAGGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 6660
Db 6930 TCTATGGCCAGCTCCTCGGCTAGCCAGCTGTCCGCTCCATCTCTCAAGGCAACTTGCACC 6989
Qy 6661 ACCCACGGCAAAGCCTATGATGTGGACATGGTGGATGCTAACCT----------GTTC 6708
Db 6990 GCCAACCATGACTCCCCTGACGCCGAGCTCATAGAGGCTAACCTCCTGTGGAGGCAGGAG 7049
Qy 6709 ATGGGGGGCGATGTGACTCGGATAGAGTCTGGGTCCAAAGTGGTCGTTCTGGACTCTCTC 6768
Db 7050 ATGGGCGGCAACATCACCAGGGTTGAGTCAGAGAACAAAGTGGTGATTCTGGACTCCTTC 7109
Qy 6769 GACCCAATGGTCGAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCAGAATACATGCTC 6828
Db 7110 GATCCGCTTGTGGCAGAGGAGGATGAGCGGGAGGTCTCCGTACCTGCAGAAATTCTGCGG 7169
Qy 6829 CCCAAGAAGAGGTTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCG 6888
Db 7170 AAGTCTCGGAGATTCGCCCGGGCCCTGCCCGTCTGGGCGCGGCCGGACTACAACCCCCCG 7229
Qy 6889 CTTGTGGAATCGTGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTC 6948
Db 7230 CTAGTAGAGACGTGGAAAAAGCCTGACTACGAACCACCTGTGGTCCATGGCTGCCCGCTA 7289
Qy 6949 CCTCCTCCTAGGAAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGT 7008
Db 7290 CCACCTCCACGGTCCCCTCCTGTGCCTCCGCCTCGGAAAAAGCGTACGGTGGTCCTCACC 7349
Qy 7009 GAGGACTCCATAGGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCC 7068
Db 7350 GAATCAACCCTATCTACTGCCTTGGCCGAGCTTGCCACCAAAAGTTTTGGCAGCTCCTCA 7409
Qy 7069 CCAAGCGGCGATTCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACG 7128
Db 7410 ACTTCCGGCATTACGGGCGACAATACGACAACATCCTCTGAGCCCGCC---------- 7457
Qy 7129 CCTCCTGATGAGTTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCTCGAG 7188
Db 7458 CCTTCTGGCTGCCCCCCCGACTCCGACGTTGAGTCCTATTCTTCCATGCCCCCCCTGGAG 7517
Qy 7189 GGGGAGCTTGGAGATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCAG 7248




```
RESULT 12
PCT-US91-02225-9
; Sequence 9, Application PC/TUS9102225
;  GENERAL INFORMATION:
;    APPLICANT:  HOUGHTON, MICHAEL
;    APPLICANT:  CHOO, QUI-LIM
;    APPLICANT:  KUO, GEORGE
;    TITLE OF INVENTION:  COMBINATIONS OF HEPATITIS C VIRUS
;    TITLE OF INVENTION:
;    TITLE OF INVENTION:  ANTIBODIES
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Morrison & Foerster
;      STREET:  545 Middlefield Road, Suite 200
;      CITY:  Menlo Park
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94025
```

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US91/02225
;      FILING DATE:  19910329
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CIOTTI, THOMAS E.
;      REGISTRATION NUMBER:  21,013
;      REFERENCE/DOCKET NUMBER:  2300-0101.44
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 327-7250
;      TELEFAX:  (415) 327-2951
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9401 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
PCT-US91-02225-9

  Query Match             45.3%;  Score 4118;  DB 5;  Length 9401;
  Best Local Similarity   66.2%;  Pred. No. 0;
  Matches 6029;  Conservative   20;  Mismatches 2971;  Indels   90;  Gaps    5;

Qy          1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 60
              |||||||| |||||||||||||||||:|||||:|||| | ||||||||||||||||||||
Db        342 ATGAGCACGAATCCTAAACCTCAAARAAAAAMCAAACGTAACACCAACCGTCGCCCACAG 401

Qy         61 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 120
              ||||| ||||| ||||| ||||| |||||||||| ||||| ||||||||||||||||||
Db        402 GACGTCAAGTTCCCGGGTGGCGGTCAGATCGTTGGTGGAGTTTACTTGTTGCCGCGCAGG 461

Qy        121 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 180
              ||||| || |||||||||||||||||| || |||||||| |||||||| || || || ||
Db        462 GGCCCTAGATTGGGTGTGCGCGCGACGAGAAAGACTTCCGAGCGGTCGCAACCTCGAGGT 521

Qy        181 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 240
              || || ||||| ||||| || | ||| || ||  |||| ||||| |||||| | || ||
Db        522 AGACGTCAGCCTATCCCCAAGGCTCGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGG 581

Qy        241 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 300
              ||||| |||||||| || || ||||||||  ||| ||||| ||||||||||||||| |||
Db        582 TACCCTTGGCCCCTCTATGGCAATGAGGGCTGCGGGTGGGCGGGATGGCTCCTGTCTCCC 641

Qy        301 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 360
              || || || || ||    |||||||||| |||||||||| |||||||||||  |||||
Db        642 CGTGGCTCTCGGCCTAGCTGGGGCCCCACAGACCCCCGGCGTAGGTCGCGCAATTTGGGT 701

Qy        361 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 420
              |||||||||||||||| |||||||||| |||||||||||||||||||| ||  ||||
Db        702 AAGGTCATCGATACCCTTACGTGCGGCTTCGCCGACCTCATGGGGTACATACCGCTCGTC 761

Qy        421 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 480
              |||||||| || || |||| ||||| || || |||||||||| | || ||||| |||
Db        762 GGCGCCCCTCTTGGAGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAAGAC 821

Qy        481 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 540
              || || || | |||||||||||||| | || |||||||| || ||||:|||| | ||||||
Db        822 GGCGTGAACTATGCAACAGGGAACCTTCCTGGTTGCTCTTTCTCTAYCTTCCTTCTGGCC 881

Qy        541 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 600
              ||||| || ||| | ||    || |  || ||  | |||||   |||  ||
Db        882 CTGCTCTCTTGCTTGACTGTGCCCGCTTCGGCCTACCAAGTGCGCAACTCCACGGGGCTT 941

Qy        601 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 660
              |||   || || || || ||| | ||    || |||   |   ||      | |||  ||
Db        942 TACCACGTCACCAATGATTGCCCTAACTCGAGTATTGTGTACGAGGCGGCCGATGCCATC 1001

Qy        661 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 720
              || |||   || |||||||||| ||||      | ||| || || ||  ||| ||| |
Db       1002 CTGCACACTCCGGGGTGCGTCCCTTGCGTTCGTGAGGGCAACGCCTCGAGGTGTTGGGTG 1061

Qy        721 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 780
              || |  | || |  ||||||      |      |  || | ||  |    | ||
Db       1062 GCGATGACCCCTACGGTGGCCACCAGGGATGGCAAACTCCCCGCGACGCAGCTTCGACGT 1121

Qy        781 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 840
              |||||||  || ||||    |  |||||| |||||  || ||||||||||||||||||||
Db       1122 CACATCGATCTGCTTGTCGGGAGCGCCACCCTCTGTTCGGCCCTCTACGTGGGGGACCTA 1181

Qy        841 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 900
              |||||    ||  | || |  | |||| ||||||   |||| ||   || |||||||
Db       1182 TGCGGGTCTGTCTTTCTTGTCGGCCAACTGTTCACCTTCTCTCCCAGGCGCCACTGGACG 1241

Qy        901 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 960
              ||||  |||||||||||| ||||| || ||    || || || ||||||||||||||||
Db       1242 ACGCAAGGTTGCAATTGCTCTATCTATCCCGGCCATATAACGGGTCACCGCATGGCATGG 1301

Qy        961 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1020
              || |||||||||||||||| || ||| |  |  || |  |||  |  | | ||  ||
Db       1302 GATATGATGATGAACTGGTCCCCTACGACGGCGTTGGTARTGGCTCAGCTGCTCCGGATC 1361

Qy       1021 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1080
              ||  | | |||  | ||||| ||    || ||||| ||||| ||| ||   ||| | ||
Db       1362 CCACAAGCCATCTTGGACATGATCGCTGGTGCTCACTGGGGAGTCCTGGCGGGCATAGCG 1421

Qy       1081 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1140
              || ||||| |||  |||    |||||||| ||| | ||  | || ||| |    |||||
Db       1422 TATTTCTCCATGGTGGGGAACTGGGCGAAGGTCCTGGTAGTGCTGCTGCTATTTGCCGGC 1481

Qy       1141 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1200
              || ||||||    |||||          ||||| |||||  || ||          | |
Db       1482 GTCGACGCGGAAACCCACGTCACCGGGGGAAGTGCCGGCCACACTGTGTCTGGATTTGTT 1541

Qy       1201 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1260
              ||| |  |  |    ||| ||| ||||||  |||||||  | |||||||| ||||| |||
Db       1542 AGCCTCCTCGCACCAGGCGCCAAGCAGAACGTCCAGCTGATCAACACCAACGGCAGTTGG 1601

Qy       1261 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1320
              ||| ||||  |||| ||||||||||||||||||    | |  |||||||||    | ||
Db       1602 CACCTCAATAGCACGGCCCTGAACTGCAATGATAGCCTCAACACCGGCTGGTTGGCAGGG 1661

Qy       1321 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1380
              || |||||   |||||  |||||||| ||||| ||||| ||  |  | ||  ||||||
Db       1662 CTTTTCTATCACCACAAGTTCAACTCTTCAGGCTGTCCTGAGAGGCTAGCCAGCTGCCGA 1721

Qy       1381 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1440
              |   |  ||    ||| |||||| |  |   |||  |    |  |
Db       1722 CCCCTTACCGATTTTGACCAGGGCTGGGGCCCTATCAGTTATGCCAACGGAAGCGGCC-- 1779

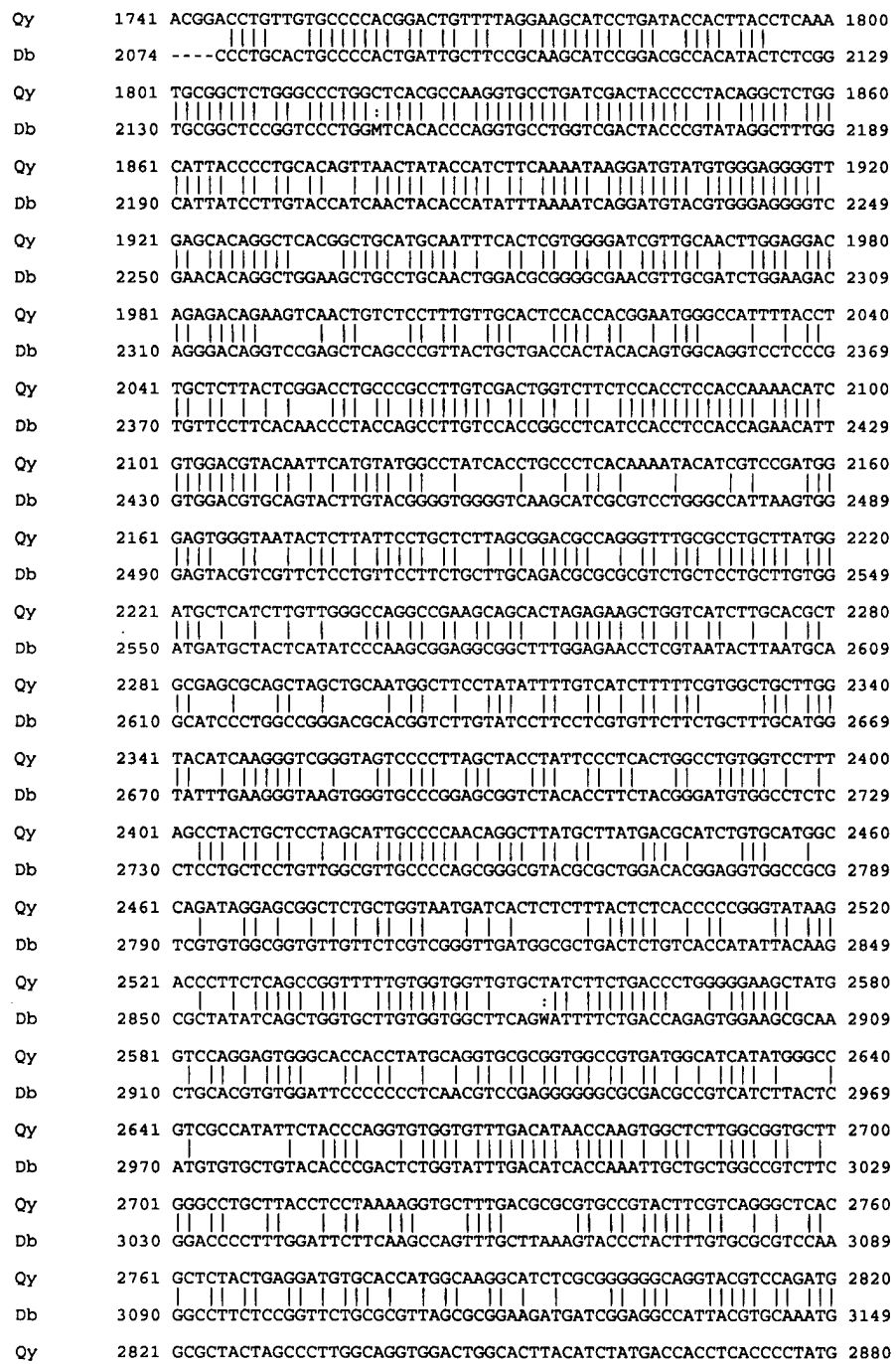
Qy       1441 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1500
                 ||   | | ||||| |||||||||||||| ||||  |  || ||  | ||  ||
Db       1780 ----CCGACCAGCGCCCCTACTGCTGGCACTACCCCCCAAAACCTTGCGGTATTGTGCCC 1835

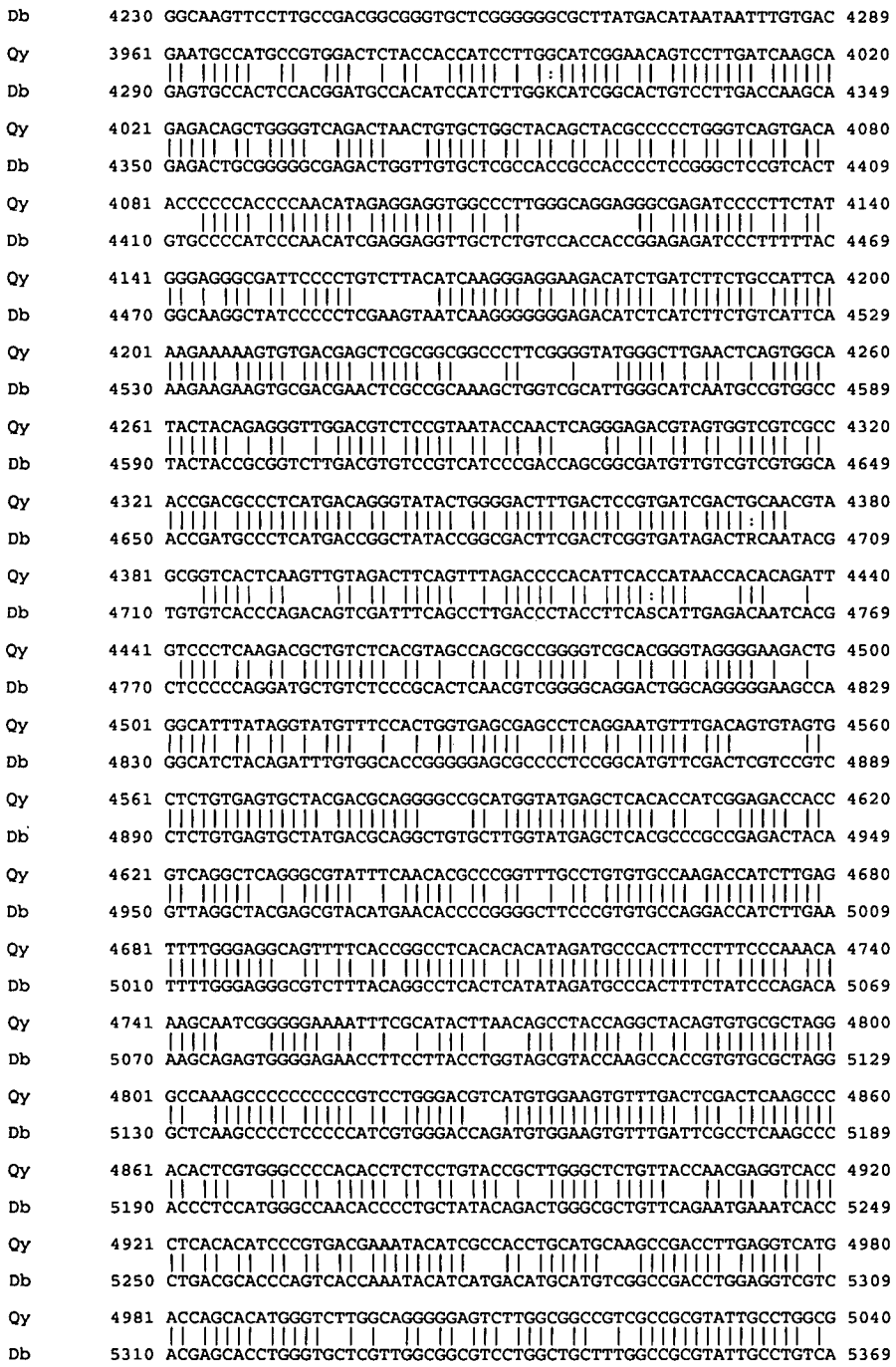
Qy       1501 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCAGCCCAGTGGTAGTGGGCACG 1560
              |||||| |||||||| || || || |||| |||||||  ||||| |||||  |||
Db       1836 GCGAAGAGTGTGTGTGGTCCGGTATATTGCTTCACTCCCAGCCCCGTGGTGGTGGGAACG 1895

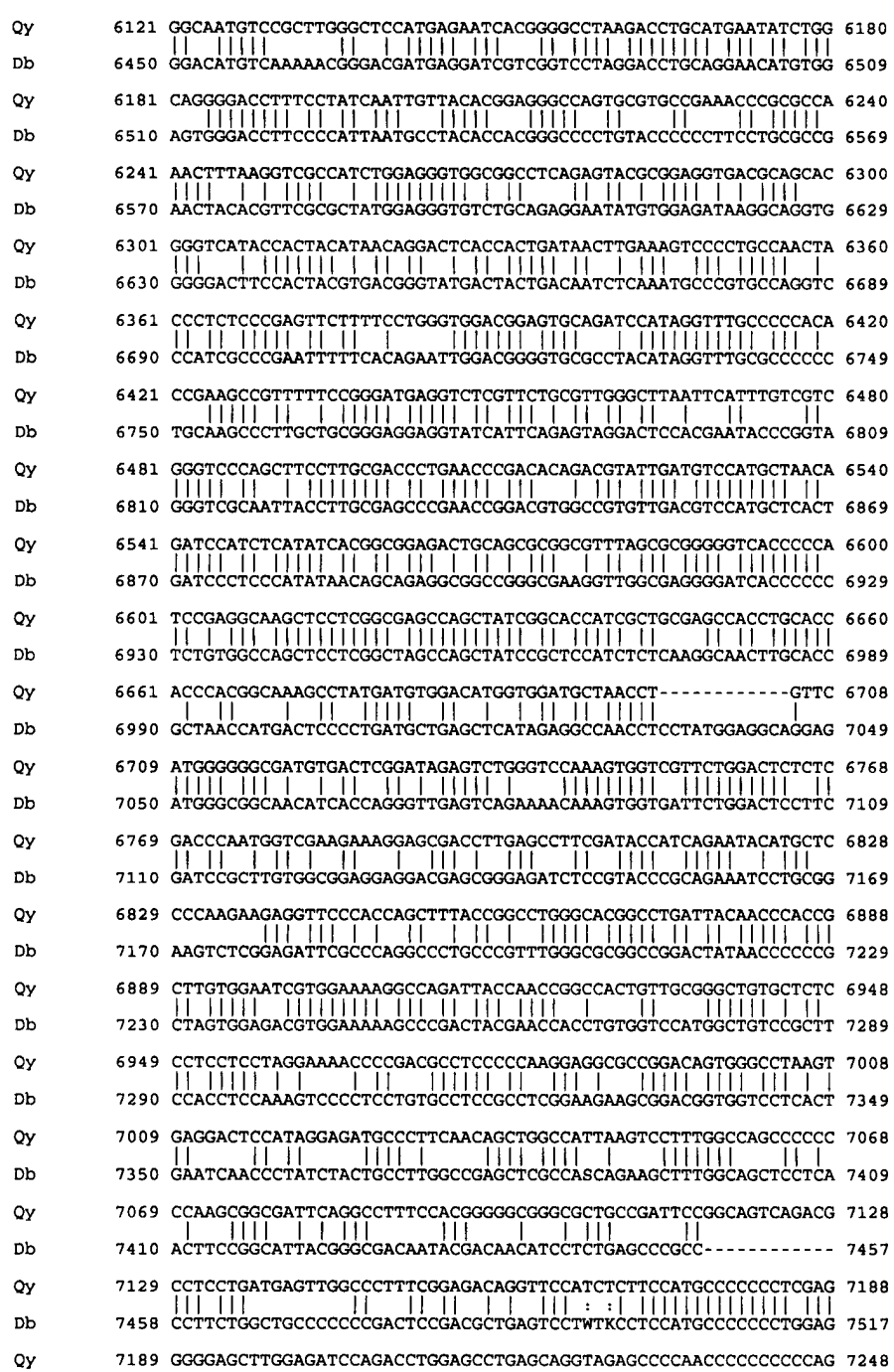
Qy       1561 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1620
              |||||||||   || |||||||| ||||  |||| || |||||  || ||  |||||| |
Db       1896 ACCGACAGGTCGGGCGCGCCCACCTACAGCTGGGGTGAAAATGATACGGACGTCTTCGTC 1955

Qy       1621 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 1680
              | ||||  ||  | ||||||||||    |||||||| || || ||||||||||  ||
Db       1956 CTTAACAATACCAGGCCACCGCTGGGCAATTGGTTCGGTTGTACCTGGATGAACTCAACT 2015

Qy       1681 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 1740
              || | ||||||    |||| || || || ||  |  ||| |   |||| ||
Db       2016 GGATTCACCAAAGTGTGCGGAGCGCCTCCTTGTGTCATCGGAGGGGCGGGCAACAACA-- 2073
```


Qy 1741 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 1800
Db 2074 ----CCCTGCACTGCCCCACTGATTGCTTCCGCAAGCATCCGGACGCCACATACTCTCGG 2129
Qy 1801 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 1860
Db 2130 TGCGGCTCCGGTCCCTGGMTCACACCCAGGTGCCTGGTCGACTACCCGTATAGGCTTTGG 2189
Qy 1861 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 1920
Db 2190 CATTATCCTTGTACCATCAACTACACCATATTTAAAATCAGGATGTACGTGGGAGGGGTC 2249
Qy 1921 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 1980
Db 2250 GAACACAGGCTGGAAGCTGCCTGCAACTGGACGCGGGGCGAACGTTGCGATCTGGAAGAC 2309
Qy 1981 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2040
Db 2310 AGGGACAGGTCCGAGCTCAGCCCGTTACTGCTGACCACTACACAGTGGCAGGTCCTCCCG 2369
Qy 2041 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2100
Db 2370 TGTTCCTTCACAACCCTACCAGCCTTGTCCACCGGCCTCATCCACCTCCACCAGAACATT 2429
Qy 2101 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2160
Db 2430 GTGGACGTGCAGTACTTGTACGGGGTGGGGTCAAGCATCGCGTCCTGGGCCATTAAGTGG 2489
Qy 2161 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2220
Db 2490 GAGTACGTCGTTCTCCTGTTCCTTCTGCTTGCAGACGCGCGCGTCTGCTCCTGCTTGTGG 2549
Qy 2221 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2280
Db 2550 ATGATGCTACTCATATCCCAAGCGGAGGCGGCTTTGGAGAACCTCGTAATACTTAATGCA 2609
Qy 2281 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATATTTTGTCATCTTTTTCGTGGCTGCTTGG 2340
Db 2610 GCATCCCTGGCCGGGACGCACGGTCTTGTATCCTTCCTCGTGTTCTTCTGCTTTGCATGG 2669
Qy 2341 TACATCAAGGGTCGGGTAGTCCCCTTAGCTACCTATTCCCTCACTGGCCTGTGGTCCTTT 2400
Db 2670 TATTTGAAGGGTAAGTGGGTGCCCGGAGCGGTCTACACCTTCTACGGGATGTGGCCTCTC 2729
Qy 2401 AGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTGCATGGC 2460
Db 2730 CTCCTGCTCCTGTTGGCGTTGCCCCAGCGGGCGTACGCGCTGGACACGGAGGTGGCCGCG 2789
Qy 2461 CAGATAGGAGCGGCTCTGCTGGTAATGATCACTCTCTTTACTCTCACCCCCGGGTATAAG 2520
Db 2790 TCGTGTGGCGGTGTTGTTCTCGTCGGGTTGATGGCGCTGACTCTGTCACCATATTACAAG 2849
Qy 2521 ACCCTTCTCAGCCGGTTTTTGTGGTGGTTGTGCTATCTTCTGACCCTGGGGGAAGCTATG 2580
Db 2850 CGCTATATCAGCTGGTGCTTGTGGTGGCTTCAGWATTTTCTGACCAGAGTGGAAGCGCAA 2909
Qy 2581 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGCATCATATGGGCC 2640
Db 2910 CTGCACGTGTGGATTCCCCCCCTCAACGTCCGAGGGGGGCGCGACGCCGTCATCTTACTC 2969
Qy 2641 GTCGCCATATTCTACCCAGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 2700
Db 2970 ATGTGTGCTGTACACCCGACTCTGGTATTTGACATCACCAAATTGCTGCTGGCCGTCTTC 3029
Qy 2701 GGGCCTGCTTACCTCCTAAAAGGTGCTTTGACGCGCGTGCCGTACTTCGTCAGGGCTCAC 2760
Db 3030 GGACCCCTTTGGATTCTTCAAGCCAGTTTGCTTAAAGTACCCTACTTTGTGCGCGTCCAA 3089
Qy 2761 GCTCTACTGAGGATGTGCACCATGGCAAGGCATCTCGCGGGGGGCAGGTACGTCCAGATG 2820
Db 3090 GGCCTTCTCCGGTTCTGCGCGTTAGCGCGGAAGATGATCGGAGGCCATTACGTGCAAATG 3149
Qy 2821 GCGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 2880
Db 3150 GTCATCATTAAGTTAGGGGCGCTTACTGGCACCTATGTTTATAACCATCTCACTCCTCTT 3209
Qy 2881 TCGGATTGGGCTGCTAGTGGCCTGCGGGACCTGGCGGTCGCCGTTGAGCCTATCATCTTC 2940
Db 3210 CGGGACTGGGCGCACAACGGCTTGCGAGATCTGGCCGTGGCTGTAGAGCCAGTCGTCTTC 3269
Qy 2941 AGTCCGATGGAGAAGAAAGTCATTGTCTGGGGAGCGGAGACAGCTGCTTGTGGGGACATT 3000
Db 3270 TCCCAAATGGAGACCAAGCTCATCACGTGGGGGGCAGATACCGCCGCGTGCGGTGACATC 3329
Qy 3001 TTACACGGACTTCCCGTGTCCGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3060
Db 3330 ATCAACGGCTTGCCTGTTTCCGCCCGCAGGGGCCGGGAGATACTGCTCGGGCCAGCCGAT 3389
Qy 3061 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3120
Db 3390 GGAATGGTCTCCAAGGGGTGGAGGTTGCTGGCGCCCATCACGGCGTACGCCCAGCAGACA 3449
Qy 3121 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCT 3180
Db 3450 AGGGGCCTCCTAGGGTGCATAATCACCAGCCTAACTGGCCGGGACAAAAACCAAGTGGAG 3509
Qy 3181 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3240
Db 3510 GGTGAGGTCCAGATTGTGTCAACTGCTGCCCAAACCTTCCTGGCAACGTGCATCAATGGG 3569
Qy 3241 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3300
Db 3570 GTGTGCTGGACTGTCTACCACGGGGCCGGAACGAGGACCATCGCGTCACCCAAGGGTCCT 3629
Qy 3301 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGG 3360
Db 3630 GTCATCCAGATGTATACCAATGTAGACCAAGACCTTGTGGGCTGGCCCGCTYCGCAAGGT 3689
Qy 3361 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3420
Db 3690 ASCCGCTCATTGACACCCTGCACTTGCGGCTCCTCGGACCTTTACCTGGTCACGAGGCAC 3749
Qy 3421 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3480
Db 3750 GCCGATGTCATTCCCGTGCGCCGGCGGGGTGATAGCAGGGGCAGCCTGCTGTCGCCCCGG 3809
Qy 3481 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3540
Db 3810 CCCATTTCCTACTTGAAAGGCTCCTCGGGGGGTCCGCTGTTGTGCCCCGCGGGGCACGCC 3869
Qy 3541 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3600
Db 3870 GTGGGCATATTTAGGGCCGCGGTGTGCACCCGTGGAGTGGCTAAGGCGGTGGACTTTATC 3929
Qy 3601 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 3660
Db 3930 CCTGTGGAGAACCTAGAGACAACCATGAGGTCCCCGGTGTTCACGGATAACTCCTCTCCA 3989
Qy 3661 CCTGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 3720
Db 3990 CCAGTAGTGCCCCAGAGCTTCCAGGTGGCTCACCTCCATGCTCCCACAGGCAGCGGCAAA 4049
Qy 3721 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 3780
Db 4050 AGCACCAAGGTCCCGGCTGCATATGCAGCTCAGGGCTATAAGGTGCTAGTACTCAACCCC 4109
Qy 3781 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 3840
Db 4110 TCTGTTGCTGCAACACTGGGCTTTGGTGCTTACATGTCCAAGGCTCATGGGATCGATCYT 4169
Qy 3841 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 3900
Db 4170 AACATCAGGACCGGGGTGAGAACAATTACCACTGGCAGCCCCATCACGTACTCCACCTAC 4229
Qy 3901 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 3960


Db 4230 GGCAAGTTCCTTGCCGACGGCGGGTGCTCGGGGGGCGCTTATGACATAATAATTTGTGAC 4289
Qy 3961 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4020
Db 4290 GAGTGCCACTCCACGGATGCCACATCCATCTTGGKCATCGGCACTGTCCTTGACCAAGCA 4349
Qy 4021 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4080
Db 4350 GAGACTGCGGGGGCGAGACTGGTTGTGCTCGCCACCGCCACCCCTCCGGGCTCCGTCACT 4409
Qy 4081 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4140
Db 4410 GTGCCCCATCCCAACATCGAGGAGGTTGCTCTGTCCACCACCGGAGAGATCCCTTTTTAC 4469
Qy 4141 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4200
Db 4470 GGCAAGGCTATCCCCCTCGAAGTAATCAAGGGGGGGAGACATCTCATCTTCTGTCATTCA 4529
Qy 4201 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4260
Db 4530 AAGAAGAAGTGCGACGAACTCGCCGCAAAGCTGGTCGCATTGGGCATCAATGCCGTGGCC 4589
Qy 4261 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4320
Db 4590 TACTACCGCGGTCTTGACGTGTCCGTCATCCCGACCAGCGGCGATGTTGTCGTCGTGGCA 4649
Qy 4321 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4380
Db 4650 ACCGATGCCCTCATGACCGGCTATACCGGCGACTTCGACTCGGTGATAGACTRCAATACG 4709
Qy 4381 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4440
Db 4710 TGTGTCACCCAGACAGTCGATTTCAGCCTTGACCCTACCTTCASCATTGAGACAATCACG 4769
Qy 4441 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4500
Db 4770 CTCCCCCAGGATGCTGTCTCCCGCACTCAACGTCGGGGCAGGACTGGCAGGGGGAAGCCA 4829
Qy 4501 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4560
Db 4830 GGCATCTACAGATTTGTGGCACCGGGGGAGCGCCCCTCCGGCATGTTCGACTCGTCCGTC 4889
Qy 4561 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4620
Db 4890 CTCTGTGAGTGCTATGACGCAGGCTGTGCTTGGTATGAGCTCACGCCCGCCGAGACTACA 4949
Qy 4621 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 4680
Db 4950 GTTAGGCTACGAGCGTACATGAACACCCCGGGGCTTCCCGTGTGCCAGGACCATCTTGAA 5009
Qy 4681 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 4740
Db 5010 TTTTGGGAGGGCGTCTTTACAGGCCTCACTCATATAGATGCCCACTTTCTATCCCAGACA 5069
Qy 4741 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 4800
Db 5070 AAGCAGAGTGGGGAGAACCTTCCTTACCTGGTAGCGTACCAAGCCACCGTGTGCGCTAGG 5129
Qy 4801 GCCAAAGCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 4860
Db 5130 GCTCAAGCCCCTCCCCCATCGTGGGACCAGATGTGGAAGTGTTTGATTCGCCTCAAGCCC 5189
Qy 4861 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 4920
Db 5190 ACCCTCCATGGGCCAACACCCCTGCTATACAGACTGGGCGCTGTTCAGAATGAAATCACC 5249
Qy 4921 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 4980
Db 5250 CTGACGCACCCAGTCACCAAATACATCATGACATGCATGTCGGCCGACCTGGAGGTCGTC 5309
Qy 4981 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5040
Db 5310 ACGAGCACCTGGGTGCTCGTTGGCGGCGTCCTGGCTGCTTTGGCCGCGTATTGCCTGTCA 5369
Qy 5041 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5100
Db 5370 ACAGGCTGCGTGGTCATAGTGGGCAGGGTCGTCTTGTCCGGGAAGCCGGCAATCATACCT 5429
Qy 5101 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5160
Db 5430 GACAGGGAAGTCCTCTACCGAGAGTTCGATGAGATGGAAGAGTGCTCTCAGCACTTACCG 5489
Qy 5161 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5220
Db 5490 TACATCGAGCAAGGGATGATGCTCGCCGAGCAGTTCAAGCAGAAGGCCCTCGGCCTCCTG 5549
Qy 5221 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5280
Db 5550 CAGACCGCGTCCCGTCAGGCAGAGGTTATCGCCCCTGCTGTCCAGACCAACTGGCAAAAA 5609
Qy 5281 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5340
Db 5610 CTCGAGACCTTCTGGGCGAAGCATATGTGGAACTTCATCAGTGGGATACAATACTTGGCG 5669
Qy 5341 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5400
Db 5670 GGCTTGTCAACGCTGCCTGGTAACCCCGCCATTGCTTCATTGATGGCTTTTACAGCTGCT 5729
Qy 5401 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5460
Db 5730 GTCACCAGCCCACTAACCACTAGCCAAACCCTCCTCTTCAACATATTGGGGGGGTGGGTG 5789
Qy 5461 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5520
Db 5790 GCTGCCCAGCTCGCCGCCCCCGGTGCCGCTACTGCCTTTGTGGGCGCTGGCTTAGCTGGC 5849
Qy 5521 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5580
Db 5850 GCCGCCATCGGCAGTGTTGGACTGGGGAAGGTCCTCATAGACATCCTTGCAGGGTATGGC 5909
Qy 5581 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5640
Db 5910 GCGGGCGTGGCGGGAGCTCTTGTGGCATTCAAGATCATGAGCGGTGAGGTCCCCTCCACG 5969
Qy 5641 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 5700
Db 5970 GRGGACCTGGTCAATCTACTGCCCGCCATCCTCTCGCCCGGAGCCCTCGTAGTCGGCGTG 6029
Qy 5701 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 5760
Db 6030 GTCTGTGCAGCAATACTGCGCCGGCACGTTGGCCCGGGCGAGGGGGCAGTGCAGTGGATG 6089
Qy 5761 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 5820
Db 6090 AACCGGCTGATAGCCTTCGCCTCCCGGGGGAACCATGTTTCCCCCACGCACTACGTGCCG 6149
Qy 5821 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 5880
Db 6150 GAGAGCGATGCAGCTGCCCGCGTCACTGCCATACWCWGCAGCCTCACTGTAACCCAGCTC 6209
Qy 5881 CTCAGAAGACTCCACAACTGGATTACTGAGGACTGCCCCATCCCATGCGGCGGCTCGTGG 5940
Db 6210 CTGAGGCGACTGCACCAGTGGATAAGCTCGGAGTGTACCACTCCATGCTCCGGTTCCTGG 6269
Qy 5941 CTCCGCGATGTGTGGGACTGGGTTTGCACCATCCTAACAGACTTTAAAAATTGGCTGACC 6000
Db 6270 CTAAGGGACATCTGGGACTGGATATGCGAGGTGTTGAGCGACTTTAAGACCTGGCTAAAA 6329
Qy 6001 TCCAAATTATTCCCAAAGATGCCCGGCCTCCCCTTTGTCTCCTGTCAAAAGGGGTACAAG 6060
Db 6330 GCTAAGCTCATGCCACAGCTGCCTGGGATCCCCTTTGTGTCCTGCCAGCGCGGGTATAAG 6389
Qy 6061 GGCGTGTGGGCCGGCACTGGCATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6120
Db 6390 GGGGTCTGGCGAGKGGACGGCATCATGCACACTCGCTGCCACTGTGGAGCTGAGATCACT 6449


Qy 6121 GGCAATGTCCGCTTGGGCTCCATGAGAATCACGGGGCCTAAGACCTGCATGAATATCTGG 6180
Db 6450 GGACATGTCAAAAACGGGACGATGAGGATCGTCGGTCCTAGGACCTGCAGGAACATGTGG 6509
Qy 6181 CAGGGGACCTTTCCTATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCGCCA 6240
Db 6510 AGTGGGACCTTCCCCATTAATGCCTACACCACGGGCCCCTGTACCCCCCTTCCTGCGCCG 6569
Qy 6241 AACTTTAAGGTCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6300
Db 6570 AACTACACGTTCGCGCTATGGAGGGTGTCTGCAGAGGAATATGTGGAGATAAGGCAGGTG 6629
Qy 6301 GGGTCATACCACTACATAACAGGACTCACCACTGATAACTTGAAAGTCCCCTGCCAACTA 6360
Db 6630 GGGGACTTCCACTACGTGACGGGTATGACTACTGACAATCTCAAATGCCCGTGCCAGGTC 6689
Qy 6361 CCCTCTCCCGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCACA 6420
Db 6690 CCATCGCCCGAATTTTTCACAGAATTGGACGGGGTGCGCCTACATAGGTTTGCGCCCCCC 6749
Qy 6421 CCGAAGCCGTTTTTCCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6480
Db 6750 TGCAAGCCCTTGCTGCGGGAGGAGGTATCATTCAGAGTAGGACTCCACGAATACCCGGTA 6809
Qy 6481 GGGTCCCAGCTTCCTTGCGACCCTGAACCCGACACAGACGTATTGATGTCCATGCTAACA 6540
Db 6810 GGGTCGCAATTACCTTGCGAGCCCGAACCGGACGTGGCCGTGTTGACGTCCATGCTCACT 6869
Qy 6541 GATCCATCTCATATCACGGCGGAGACTGCAGCGCGGCGTTTAGCGCGGGGGTCACCCCCA 6600
Db 6870 GATCCCTCCCATATAACAGCAGAGGCGGCCGGGCGAAGGTTGGCGAGGGGATCACCCCCC 6929
Qy 6601 TCCGAGGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 6660
Db 6930 TCTGTGGCCAGCTCCTCGGCTAGCCAGCTATCCGCTCCATCTCTCAAGGCAACTTGCACC 6989
Qy 6661 ACCCACGGCAAAGCCTATGATGTGGACATGGTGGATGCTAACCT----------GTTC 6708
Db 6990 GCTAACCATGACTCCCCTGATGCTGAGCTCATAGAGGCCAACCTCCTATGGAGGCAGGAG 7049
Qy 6709 ATGGGGGGCGATGTGACTCGGATAGAGTCTGGGTCCAAAGTGGTCGTTCTGGACTCTCTC 6768
Db 7050 ATGGGCGGCAACATCACCAGGGTTGAGTCAGAAAACAAAGTGGTGATTCTGGACTCCTTC 7109
Qy 6769 GACCCAATGGTCGAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCAGAATACATGCTC 6828
Db 7110 GATCCGCTTGTGGCGGAGGAGGACGAGCGGGAGATCTCCGTACCCGCAGAAATCCTGCGG 7169
Qy 6829 CCCAAGAAGAGGTTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCG 6888
Db 7170 AAGTCTCGGAGATTCGCCCAGGCCCTGCCCGTTTGGGCGCGGCCGGACTATAACCCCCCG 7229
Qy 6889 CTTGTGGAATCGTGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTC 6948
Db 7230 CTAGTGGAGACGTGGAAAAAGCCCGACTACGAACCACCTGTGGTCCATGGCTGTCCGCTT 7289
Qy 6949 CCTCCTCCTAGGAAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGT 7008
Db 7290 CCACCTCCAAAGTCCCCTCCTGTGCCTCCGCCTCGGAAGAAGCGGACGGTGGTCCTCACT 7349
Qy 7009 GAGGACTCCATAGGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCC 7068
Db 7350 GAATCAACCCTATCTACTGCCTTGGCCGAGCTCGCCASCAGAAGCTTTGGCAGCTCCTCA 7409
Qy 7069 CCAAGCGGCGATTCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACG 7128
Db 7410 ACTTCCGGCATTACGGGCGACAATACGACAACATCCTCTGAGCCCGCC---------- 7457
Qy 7129 CCTCCTGATGAGTTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCTCGAG 7188
Db 7458 CCTTCTGGCTGCCCCCCCGACTCCGACGCTGAGTCCTWTKCCTCCATGCCCCCCCTGGAG 7517
Qy 7189 GGGGAGCTTGGAGATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCAG 7248
Db 7518 GGGGAGCCTGGGGATCCGGA---------- 7537
Qy 7249 GGGGGGGTGGCAGCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGAC 7308
Db 7538 ----------TCTTAGCGACGGGTCATGGTCAACGGTCAGTAGTGAGGCCAACGCG 7583
Qy 7309 GACTCCGTCGTGTGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGT 7368
Db 7584 GAGGATGTCGTGTGCTGCTCAATGTCTTACTCTTGGACAGGCGCACTCGTCACCCCGTGC 7643
Qy 7369 AGTCCCGAAGAGGAGAAGTTACCGATTAACCCCTTGAGCAACTCCCTGTTGCGATATCAC 7428
Db 7644 GCCGCGGAAGAACAGAAACTGCCCATCAATGCACTAAGCAACTCGTTGCTACGTCACCAC 7703
Qy 7429 AACAAGGTGTACTGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTT 7488
Db 7704 AATTTGGTGTATTCCACCACCTCACGCAGTGCTTGCCAAAGGCAGAAGAAAGTCACATTT 7763
Qy 7489 GATAGGATGCAAGTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCG 7548
Db 7764 GACAGACTGCAAGTTCTGGACAGCCATTACCAGGACGTACTCAAGGAGGTTAAAGCAGCG 7823
Qy 7549 GCCTCCAAGGTCACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCC 7608
Db 7824 GCGTCAAAAGTGAAGGCTAACTTBCTATCCGTAGAGGAAGCTTGCAGCCTGACGCCCCCA 7883
Qy 7609 CATTCTGCAAGATCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGG 7668
Db 7884 CACTCAGCCAAATCCAAGTTTGGTTATGGGGCAAAAGACGTCCGTTGCCATGCCAGAAAG 7943
Qy 7669 GCCGTTAACCACATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATT 7728
Db 7944 GCCGTAACCCACATCAACTCCGTGTGGAAAGACCTTCTGGAAGACAATGTAACACCAATA 8003
Qy 7729 CCCACAACCATTATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAG 7788
Db 8004 GACACTACCATCATGGCTAAGAACGAGGTTTTCTGCGTTCAGCCTGAGAAGGGGGGTCGT 8063
Qy 7789 AAAGCAGCTCGCCTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCC 7848
Db 8064 AAGCCAGCTCGTCTCATCGTGTTCCCCGATCTGGGCGTGCGCGTGTGCGAAAAGATGGCT 8123
Qy 7849 CTTTATGACATTACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAG 7908
Db 8124 TTGTACGACGTGGTTACAAAGCTCCCCTTGGCCGTGATGGGAAGCTCCTACGGATTCCAA 8183
Qy 7909 TATTCCCCCGCTCAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCT 7968
Db 8184 TACTCACCAGGACAGCGGGTTGAATTCCTCGTGCAAGCGTGGAAGTCCAAGAAAACCCCA 8243
Qy 7969 ATGGGTTTTTCGTATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGG 8028
Db 8244 ATGGGGTTCTCGTATGATACCCGCTGCTTTGACTCCACAGTCACTGAGAGCGACATCCGT 8303
Qy 8029 ACTGAGGAGTCCATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACAC 8088
Db 8304 ACGGAGGAGGCAATCTACCAATGTTGTGACCTCGACCCCCAAGCCCGCGTGGCCATCAAG 8363
Qy 8089 TCGCTAACTGAGAGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGC 8148
Db 8364 TCCCTCACCGAGAGGCTTTATGTTGGGGGCCCTCTTACCAATTCARGGGGGGAGAACTGC 8423
Qy 8149 GGGTACAGGCGTTGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACA 8208
Db 8424 GGCTATCGCAGGTGCCGCGCGAGCGGCGTACTGACAACTAGCTGTGGTAACACCCTCACT 8483
Qy 8209 TGCTACGTGAAAGCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTG 8268
Db 8484 TGCTACATCAAGGCCCGGGCAGCCTGTCGAGCCGCAGGGCTCCAGGACTGCACCATGCTC 8543
Qy 8269 GTATGCGGCGATGACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGG 8328

```
Db       8544 GTGTGTGGCGACGACTTAGTCGTTATCTGTGAAAGCGCGGGGGTCCAGGAGGACGCGGCG 8603

Qy       8329 AACCTGAGAGCCTTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCC 8388

Db       8604 AGCCTGAGAGCCTTCACGGAGGCTATGACCAGGTACTCCGCCCCCCCTGGGGACCCCCCA 8663

Qy       8389 AGACCGGAGTATGATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTG 8448

Db       8664 CAACCAGAATACGACTTGGAGCTCATAACATCATGCTCCTCCAACGTGTCAGTCGCCCAC 8723

Qy       8449 GGCCCACAAGGCCGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGG 8508

Db       8724 GACGGCGCTGGAAAGAGGGTCTACTACCTCACCCGTGACCCTACAACCCCCCTCGCGAGA 8783

Qy       8509 GCTGCCTGGGAAACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAG 8568

Db       8784 GCTGCGTGGGAGACAGCAAGACACACTCCAGTCAATTCCTGGCTAGGCAACATAATCATG 8843

Qy       8569 TACGCCCCGACCATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATG 8628

Db       8844 TTTGCCCCCACACTGTGGGCGAGGATGATACTGATGACCCATTTCTTTAGCGTCCTTATA 8903

Qy       8629 GCTCAAGACACGCTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTG 8688

Db       8904 GCCAGGGACCAGCTTGAACAGGCCCTCGATTGCGAGATCTACGGGGCCTGCTACTCCATA 8963

Qy       8689 AGTCCCTTGGACCTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTG 8748

Db       8964 GAACCACTTGATCTACCTCCAATCATTCAAAGACTCCATGGCCTCAGCGCATTTTCACTC 9023

Qy       8749 CACACATACACTCCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCG 8808

Db       9024 CACAGTTACTCTCCAGGTGAAATTAATAGGGTGGCCGCATGCCTCAGAAAACTTGGGGTA 9083

Qy       8809 CCACCCCTCAGAGCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGT 8868

Db       9084 CCGCCCTTGCGAGCTTGGRGACACCGGGCCCGGAGCGTCCGCGCTAGGCTTCTGGCCAGA 9143

Qy       8869 GGGGGGAGAGCGGCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTC 8928

Db       9144 GGAGGCAGGGCTGCCATATGTGGCAAGTACCTCTTCAACTGGGCAGTAAGAACAAAGCTC 9203

Qy       8929 AAACTCACTCCATTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGC 8988

Db       9204 AAACTCACTCCAATAGCGGCCGCTGGCCAGCTGGACTTGTCCGGCTGGTTCACGGCTGGC 9263

Qy       8989 GCCGGCGGGGGCGACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTT 9048

Db       9264 TACAGCGGGGGAGACATTTATCACAGCGTGTCTCATGCCCGGCCCCGCTGGATCTGGTTT 9323

Qy       9049 GGCCTACTCCTACTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCG 9098

Db       9324 TGCCYACTCCTGCTTGCTGCAGGGGTAGGCATCTACCTCCTCCCCAACCG 9373


RESULT 13
US-09-388-874-1
; Sequence 1, Application US/09388874
; Patent No. 6284249
; GENERAL INFORMATION:
;  APPLICANT: Veronique Barban
;  TITLE OF INVENTION: VACCINE COMPOSITION FOR PREVENTING OR
;  TITLE OF INVENTION: TREATING C HEPATITIS
;  FILE REFERENCE: PMCF97-03A
;  CURRENT APPLICATION NUMBER: US/09/388,874
;  CURRENT FILING DATE: 1999-09-02
;  EARLIER APPLICATION NUMBER: PCT/FR98/00448
;  EARLIER FILING DATE: 1998-03-06
;  EARLIER APPLICATION NUMBER: 97/02,887
;  EARLIER FILING DATE: 1997-03-06
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;  LENGTH: 9379
;  TYPE: DNA
;  ORGANISM: Virus
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (320)...(9352)
US-09-388-874-1

  Query Match             45.2%;  Score 4115.2;  DB 3;  Length 9379;
  Best Local Similarity   66.3%;  Pred. No. 0;
  Matches 6042;  Conservative    0;  Mismatches 2978;  Indels   90;  Gaps    5;

Qy          1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 60

Db        320 ATGAGCACGAATCCTAAACCTCAAAAAAAAAACAAACGTAACACCAACCGTCGCCCACAG 379

Qy         61 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 120

Db        380 GACGTCAAGTTCCCGGGTGGCGGTCAGATCGTTGGTGGAGTTTACTTGTTGCCGCGCAGG 439

Qy        121 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 180

Db        440 GGCCCTAGATTGGGTGTGCGCGCGACGAGAAAGACTTCCGAGCGGTCGCAACCTCGAGGT 499

Qy        181 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 240

Db        500 AGACGTCAGCCTATCCCCAAGGCTCGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGG 559

Qy        241 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 300

Db        560 TACCCTTGGCCCCTCTATGGCAATGAGGGCTGCGGGTGGGCGGGATGGCTCCTGTCTCCC 619

Qy        301 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 360

Db        620 CGTGGCTCTCGGCCTAGCTGGGGCCCCACAGACCCCCGGCGTAGGTCGCGCAATTTGGGT 679

Qy        361 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 420

Db        680 AAGGTCATCGATACCCTTACGTGCGGCTTCGCCGACCTCATGGGGTACATACCGCTCGTC 739

Qy        421 GGCGCCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 480

Db        740 GGCGCCCCTCTTGGAGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAAGAC 799

Qy        481 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 540

Db        800 GGCGTGAACTATGCAACAGGGAACCTTCCTGGTTGCTCTTTCTCTATCTTCCTTCTGGCC 859

Qy        541 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 600

Db        860 CTGCTCTCTTGCTTGACTGTGCCCGCTTCGGCCTACCAAGTGCGCAACTCCACGGGGCTT 919

Qy        601 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 660

Db        920 TACCACGTCACCAATGATTGCCCTAACTCGAGTATTGTGTACGAGGCGGCCGATGCCATC 979

Qy        661 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 720

Db        980 CTGCACACTCCGGGGTGCGTCCCTTGCGTTCGTGAGGGCAACGCCTCGAGGTGTTGGGTG 1039

Qy        721 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 780

Db       1040 GCGATGACCCCTACGGTGGCCACCAGGGATGGCAAACTCCCCGCGACGCAGCTTCGACGT 1099

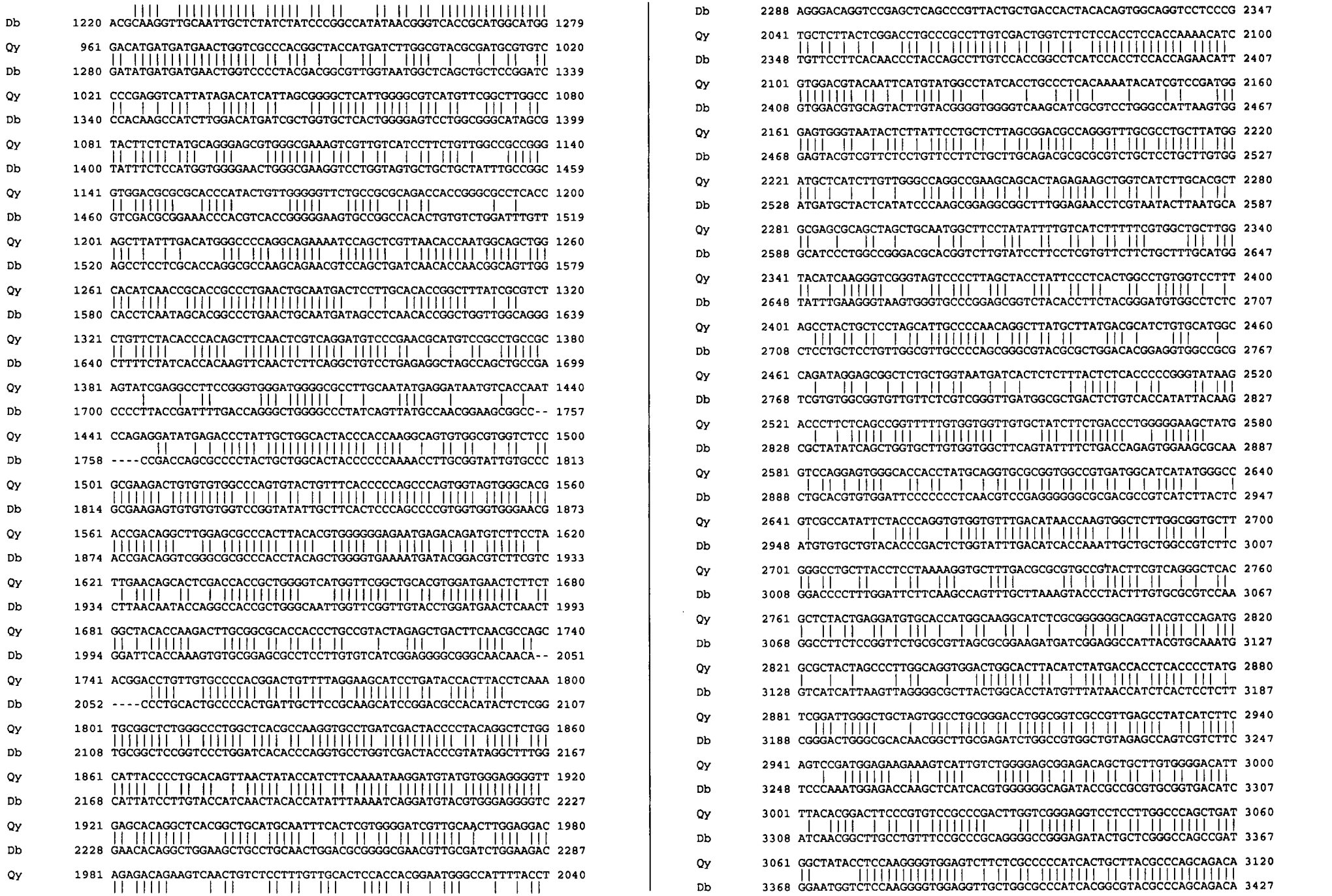
Qy        781 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 840

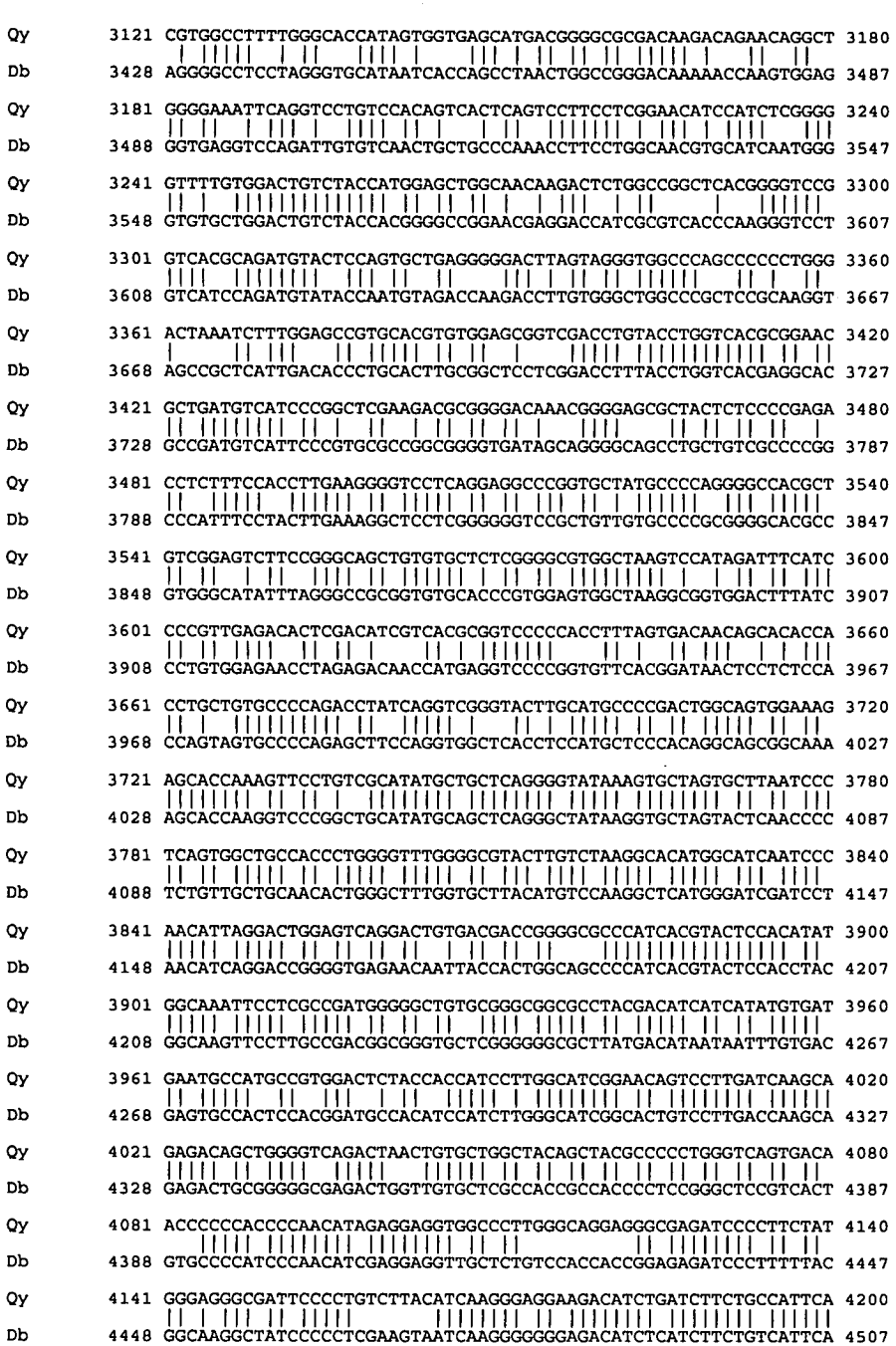
Db       1100 CACATCGATCTGCTTGTCGGGAGCGCCACCCTCTGTTCGGCCCTCTACGTGGGGGACCTA 1159

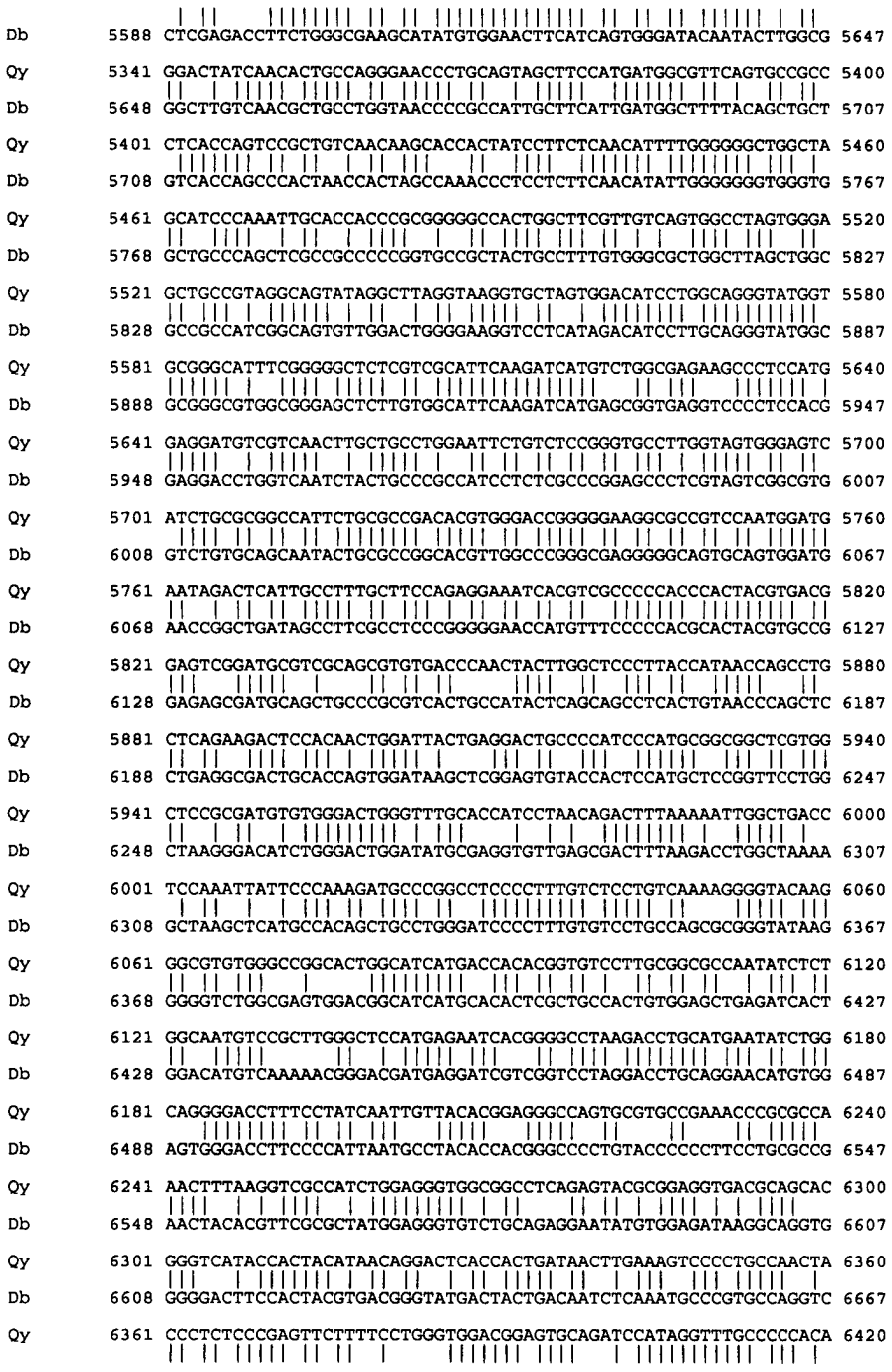
Qy        841 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 900

Db       1160 TGCGGGTCTGTCTTTCTTGTCGGCCAACTGTTCACCTTCTCTCCCAGGCGCCACTGGACG 1219

Qy        901 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 960
```




Qy 3121 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCT 3180
Db 3428 AGGGGCCTCCTAGGGTGCATAATCACCAGCCTAACTGGCCGGGACAAAAACCAAGTGGAG 3487
Qy 3181 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3240
Db 3488 GGTGAGGTCCAGATTGTGTCAACTGCTGCCCAAACCTTCCTGGCAACGTGCATCAATGGG 3547
Qy 3241 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3300
Db 3548 GTGTGCTGGACTGTCTACCACGGGGCCGGAACGAGGACCATCGCGTCACCCAAGGGTCCT 3607
Qy 3301 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGG 3360
Db 3608 GTCATCCAGATGTATACCAATGTAGACCAAGACCTTGTGGGCTGGCCCGCTCCGCAAGGT 3667
Qy 3361 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3420
Db 3668 AGCCGCTCATTGACACCCTGCACTTGCGGCTCCTCGGACCTTTACCTGGTCACGAGGCAC 3727
Qy 3421 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3480
Db 3728 GCCGATGTCATTCCCGTGCGCCGGCGGGGTGATAGCAGGGGCAGCCTGCTGTCGCCCCGG 3787
Qy 3481 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3540
Db 3788 CCCATTTCCTACTTGAAAGGCTCCTCGGGGGGTCCGCTGTTGTGCCCCGCGGGGCACGCC 3847
Qy 3541 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3600
Db 3848 GTGGGCATATTTAGGGCCGCGGTGTGCACCCGTGGAGTGGCTAAGGCGGTGGACTTTATC 3907
Qy 3601 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 3660
Db 3908 CCTGTGGAGAACCTAGAGACAACCATGAGGTCCCCGGTGTTCACGGATAACTCCTCTCCA 3967
Qy 3661 CCTGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 3720
Db 3968 CCAGTAGTGCCCCAGAGCTTCCAGGTGGCTCACCTCCATGCTCCCACAGGCAGCGGCAAA 4027
Qy 3721 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 3780
Db 4028 AGCACCAAGGTCCCGGCTGCATATGCAGCTCAGGGCTATAAGGTGCTAGTACTCAACCCC 4087
Qy 3781 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 3840
Db 4088 TCTGTTGCTGCAACACTGGGCTTTGGTGCTTACATGTCCAAGGCTCATGGGATCGATCCT 4147
Qy 3841 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 3900
Db 4148 AACATCAGGACCGGGGTGAGAACAATTACCACTGGCAGCCCCATCACGTACTCCACCTAC 4207
Qy 3901 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 3960
Db 4208 GGCAAGTTCCTTGCCGACGGCGGGTGCTCGGGGGGCGCTTATGACATAATAATTTGTGAC 4267
Qy 3961 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4020
Db 4268 GAGTGCCACTCCACGGATGCCACATCCATCTTGGGCATCGGCACTGTCCTTGACCAAGCA 4327
Qy 4021 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4080
Db 4328 GAGACTGCGGGGGCGAGACTGGTTGTGCTCGCCACCGCCACCCCTCCGGGCTCCGTCACT 4387
Qy 4081 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4140
Db 4388 GTGCCCCATCCCAACATCGAGGAGGTTGCTCTGTCCACCACCGGAGAGATCCCTTTTTAC 4447
Qy 4141 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4200
Db 4448 GGCAAGGCTATCCCCCTCGAAGTAATCAAGGGGGGGAGACATCTCATCTTCTGTCATTCA 4507
Qy 4201 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4260
Db 4508 AAGAAGAAGTGCGACGAACTCGCCGCAAAGCTGGTCGCATTGGGCATCAATGCCGTGGCC 4567
Qy 4261 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4320
Db 4568 TACTACCGCGGTCTTGACGTGTCCGTCATCCCGACCAGCGGCGATGTTGTCGTCGTGGCA 4627
Qy 4321 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4380
Db 4628 ACCGATGCCCTCATGACCGGCTATACCGGCGACTTCGACTCGGTGATAGACTGCAATACG 4687
Qy 4381 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4440
Db 4688 TGTGTCACCCAGACAGTCGATTTCAGCCTTGACCCTACCTTCACCATTGAGACAATCACG 4747
Qy 4441 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4500
Db 4748 CTCCCCCAGGATGCTGTCTCCCGCACTCAACGTCGGGGCAGGACTGGCAGGGGGAAGCCA 4807
Qy 4501 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4560
Db 4808 GGCATCTACAGATTTGTGGCACCGGGGGAGCGCCCCTCCGGCATGTTCGACTCGTCCGTC 4867
Qy 4561 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4620
Db 4868 CTCTGTGAGTGCTATGACGCAGGCTGTGCTTGGTATGAGCTCACGCCCGCCGAGACTACA 4927
Qy 4621 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 4680
Db 4928 GTTAGGCTACGAGCGTACATGAACACCCCGGGGCTTCCCGTGTGCCAGGACCATCTTGAA 4987
Qy 4681 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 4740
Db 4988 TTTTGGGAGGGCGTCTTTACAGGCCTCACTCATATAGATGCCCACTTTCTATCCCAGACA 5047
Qy 4741 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 4800
Db 5048 AAGCAGAGTGGGGAGAACCTTCCTTACCTGGTAGCGTACCAAGCCACCGTGTGCGCTAGG 5107
Qy 4801 GCCAAAGCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 4860
Db 5108 GCTCAAGCCCCTCCCCCATCGTGGGACCAGATGTGGAAGTGTTTGATTCGCCTCAAGCCC 5167
Qy 4861 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 4920
Db 5168 ACCCTCCATGGGCCAACACCCCTGCTATACAGACTGGGCGCTGTTCAGAATGAAATCACC 5227
Qy 4921 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 4980
Db 5228 CTGACGCACCCAGTCACCAAATACATCATGACATGCATGTCGGCCGACCTGGAGGTCGTC 5287
Qy 4981 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5040
Db 5288 ACGAGCACCTGGGTGCTCGTTGGCGGCGTCCTGGCTGCTTTGGCCGCGTATTGCCTGTCA 5347
Qy 5041 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5100
Db 5348 ACAGGCTGCGTGGTCATAGTGGGCAGGGTCGTCTTGTCCGGGAAGCCGGCAATCATACCT 5407
Qy 5101 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5160
Db 5408 GACAGGGAAGTCCTCTACCGAGAGTTCGATGAGATGGAAGAGTGCTCTCAGCACTTACCG 5467
Qy 5161 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5220
Db 5468 TACATCGAGCAAGGGATGATGCTCGCCGAGCAGTTCAAGCAGAAGGCCCTCGGCCTCCTG 5527
Qy 5221 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5280
Db 5528 CAGACCGCGTCCCGTCAGGCAGAGGTTATCGCCCCTGCTGTCCAGACCAACTGGCAAAAA 5587
Qy 5281 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5340


Db 5588 CTCGAGACCTTCTGGGCGAAGCATATGTGGAACTTCATCAGTGGGATACAATACTTGGCG 5647
Qy 5341 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5400
Db 5648 GGCTTGTCAACGCTGCCTGGTAACCCCGCCATTGCTTCATTGATGGCTTTTACAGCTGCT 5707
Qy 5401 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5460
Db 5708 GTCACCAGCCCACTAACCACTAGCCAAACCCTCCTCTTCAACATATTGGGGGGGTGGGTG 5767
Qy 5461 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5520
Db 5768 GCTGCCCAGCTCGCCGCCCCCGGTGCCGCTACTGCCTTTGTGGGCGCTGGCTTAGCTGGC 5827
Qy 5521 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5580
Db 5828 GCCGCCATCGGCAGTGTTGGACTGGGGAAGGTCCTCATAGACATCCTTGCAGGGTATGGC 5887
Qy 5581 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5640
Db 5888 GCGGGCGTGGCGGGAGCTCTTGTGGCATTCAAGATCATGAGCGGTGAGGTCCCCTCCACG 5947
Qy 5641 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 5700
Db 5948 GAGGACCTGGTCAATCTACTGCCCGCCATCCTCTCGCCCGGAGCCCTCGTAGTCGGCGTG 6007
Qy 5701 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 5760
Db 6008 GTCTGTGCAGCAATACTGCGCCGGCACGTTGGCCCGGGCGAGGGGGCAGTGCAGTGGATG 6067
Qy 5761 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 5820
Db 6068 AACCGGCTGATAGCCTTCGCCTCCCGGGGGAACCATGTTTCCCCCACGCACTACGTGCCG 6127
Qy 5821 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 5880
Db 6128 GAGAGCGATGCAGCTGCCCGCGTCACTGCCATACTCAGCAGCCTCACTGTAACCCAGCTC 6187
Qy 5881 CTCAGAAGACTCCACAACTGGATTACTGAGGACTGCCCCATCCCATGCGGCGGCTCGTGG 5940
Db 6188 CTGAGGCGACTGCACCAGTGGATAAGCTCGGAGTGTACCACTCCATGCTCCGGTTCCTGG 6247
Qy 5941 CTCCGCGATGTGTGGGACTGGGTTTGCACCATCCTAACAGACTTTAAAAATTGGCTGACC 6000
Db 6248 CTAAGGGACATCTGGGACTGGATATGCGAGGTGTTGAGCGACTTTAAGACCTGGCTAAAA 6307
Qy 6001 TCCAAATTATTCCCAAAGATGCCCGGCCTCCCCTTTGTCTCCTGTCAAAAGGGGTACAAG 6060
Db 6308 GCTAAGCTCATGCCACAGCTGCCTGGGATCCCCTTTGTGTCCTGCCAGCGCGGGTATAAG 6367
Qy 6061 GGCGTGTGGGCCGGCACTGGCATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6120
Db 6368 GGGGTCTGGCGAGTGGACGGCATCATGCACACTCGCTGCCACTGTGGAGCTGAGATCACT 6427
Qy 6121 GGCAATGTCCGCTTGGGCTCCATGAGAATCACGGGGCCTAAGACCTGCATGAATATCTGG 6180
Db 6428 GGACATGTCAAAAACGGGACGATGAGGATCGTCGGTCCTAGGACCTGCAGGAACATGTGG 6487
Qy 6181 CAGGGGACCTTTCCTATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCGCCA 6240
Db 6488 AGTGGGACCTTCCCCATTAATGCCTACACCACGGGCCCCTGTACCCCCCTTCCTGCGCCG 6547
Qy 6241 AACTTTAAGGTCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6300
Db 6548 AACTACACGTTCGCGCTATGGAGGGTGTCTGCAGAGGAATATGTGGAGATAAGGCAGGTG 6607
Qy 6301 GGGTCATACCACTACATAACAGGACTCACCACTGATAACTTGAAAGTCCCCTGCCAACTA 6360
Db 6608 GGGGACTTCCACTACGTGACGGGTATGACTACTGACAATCTCAAATGCCCGTGCCAGGTC 6667
Qy 6361 CCCTCTCCCGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCACA 6420
Db 6668 CCATCGCCCGAATTTTTCACAGAATTGGACGGGGTGCGCCTACATAGGTTTGCGCCCCCC 6727
Qy 6421 CCGAAGCCGTTTTTCCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6480
Db 6728 TGCAAGCCCTTGCTGCGGGAGGAGGTATCATTCAGAGTAGGACTCCACGAATACCCGGTA 6787
Qy 6481 GGGTCCCAGCTTCCTTGCGACCCTGAACCCGACACAGACGTATTGATGTCCATGCTAACA 6540
Db 6788 GGGTCGCAATTACCTTGCGAGCCCGAACCGGACGTGGCCGTGTTGACGTCCATGCTCACT 6847
Qy 6541 GATCCATCTCATATCACGGCGGAGACTGCAGCGCGGCGTTTAGCGCGGGGGTCACCCCCA 6600
Db 6848 GATCCCTCCCATATAACAGCAGAGGCGGCCGGGCGAAGGTTGGCGAGGGGATCACCCCCC 6907
Qy 6601 TCCGAGGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 6660
Db 6908 TCTGTGGCCAGCTCCTCGGCTAGCCAGCTATCCGCTCCATCTCTCAAGGCAACTTGCACC 6967
Qy 6661 ACCCACGGCAAAGCCTATGATGTGGACATGGTGGATGCTAACCT----------GTTC 6708
Db 6968 GCTAACCATGACTCCCCTGATGCTGAGCTCATAGAGGCCAACCTCCTATGGAGGCAGGAG 7027
Qy 6709 ATGGGGGGCGATGTGACTCGGATAGAGTCTGGGTCCAAAGTGGTCGTTCTGGACTCTCTC 6768
Db 7028 ATGGGCGGCAACATCACCAGGGTTGAGTCAGAAAACAAAGTGGTGATTCTGGACTCCTTC 7087
Qy 6769 GACCCAATGGTCGAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCAGAATACATGCTC 6828
Db 7088 GATCCGCTTGTGGCGGAGGAGGACGAGCGGGAGATCTCCGTACCCGCAGAAATCCTGCGG 7147
Qy 6829 CCCAAGAAGAGGTTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCG 6888
Db 7148 AAGTCTCGGAGATTCGCCCAGGCCCTGCCCGTTTGGGCGCGGCCGGACTATAACCCCCCG 7207
Qy 6889 CTTGTGGAATCGTGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTC 6948
Db 7208 CTAGTGGAGACGTGGAAAAAGCCCGACTACGAACCACCTGTGGTCCATGGCTGTCCGCTT 7267
Qy 6949 CCTCCTCCTAGGAAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGT 7008
Db 7268 CCACCTCCAAAGTCCCCTCCTGTGCCTCCGCCTCGGAAGAAGCGGACGGTGGTCCTCACT 7327
Qy 7009 GAGGACTCCATAGGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCC 7068
Db 7328 GAATCAACCCTATCTACTGCCTTGGCCGAGCTCGCCACCAGAAGCTTTGGCAGCTCCTCA 7387
Qy 7069 CCAAGCGGCGATTCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACG 7128
Db 7388 ACTTCCGGCATTACGGGCGACAATACGACAACATCCTCTGAGCCCGCC---------- 7435
Qy 7129 CCTCCTGATGAGTTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCTCGAG 7188
Db 7436 CCTTCTGGCTGCCCCCCCGACTCCGACGCTGAGTCCTATTCCTCCATGCCCCCCCTGGAG 7495
Qy 7189 GGGGAGCTTGGAGATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCAG 7248
Db 7496 GGGGAGCCTGGGGATCCGGA---------- 7515
Qy 7249 GGGGGGGTGGCAGCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGAC 7308
Db 7516 ----------TCTTAGCGACGGGTCATGGTCAACGGTCAGTAGTGAGGCCAACGCG 7561
Qy 7309 GACTCCGTCGTGTGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGT 7368
Db 7562 GAGGATGTCGTGTGCTGCTCAATGTCTTACTCTTGGACAGGCGCACTCGTCACCCCGTGC 7621
Qy 7369 AGTCCCGAAGAGGAGAAGTTACCGATTAACCCCTTGAGCAACTCCCTGTTGCGATATCAC 7428
Db 7622 GCCGCGGAAGAACAGAAACTGCCCATCAATGCACTAAGCAACTCGTTGCTACGTCACCAC 7681
Qy 7429 AACAAGGTGTACTGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTT 7488
Db 7682 AATTTGGTGTATTCCACCACCTCACGCAGTGCTTGCCAAAGGCAGAAGAAAGTCACATTT 7741

```
Qy       7489 GATAGGATGCAAGTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCG 7548
Db       7742 GACAGACTGCAAGTTCTGGACAGCCATTACCAGGACGTACTCAAGGAGGTTAAAGCAGCG 7801

Qy       7549 GCCTCCAAGGTCACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCC 7608
Db       7802 GCGTCAAAAGTGAAGGCTAACTTGCTATCCGTAGAGGAAGCTTGCAGCCTGACGCCCCCA 7861

Qy       7609 CATTCTGCAAGATCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGG 7668
Db       7862 CACTCAGCCAAATCCAAGTTTGGTTATGGGGCAAAAGACGTCCGTTGCCATGCCAGAAAG 7921

Qy       7669 GCCGTTAACCACATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATT 7728
Db       7922 GCCGTAACCCACATCAACTCCGTGTGGAAAGACCTTCTGGAAGACAATGTAACACCAATA 7981

Qy       7729 CCCACAACCATTATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAG 7788
Db       7982 GACACTACCATCATGGCTAAGAACGAGGTTTTCTGCGTTCAGCCTGAGAAGGGGGGTCGT 8041

Qy       7789 AAAGCAGCTCGCCTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCC 7848
Db       8042 AAGCCAGCTCGTCTCATCGTGTTCCCCGATCTGGGCGTGCGCGTGTGCGAAAAGATGGCT 8101

Qy       7849 CTTTATGACATTACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAG 7908
Db       8102 TTGTACGACGTGGTTACAAAGCTCCCCTTGGCCGTGATGGGAAGCTCCTACGGATTCCAA 8161

Qy       7909 TATTCCCCCGCTCAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCT 7968
Db       8162 TACTCACCAGGACAGCGGGTTGAATTCCTCGTGCAAGCGTGGAAGTCCAAGAAAACCCCA 8221

Qy       7969 ATGGGTTTTTCGTATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGG 8028
Db       8222 ATGGGGTTCTCGTATGATACCCGCTGCTTTGACTCCACAGTCACTGAGAGCGACATCCGT 8281

Qy       8029 ACTGAGGAGTCCATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACAC 8088
Db       8282 ACGGAGGAGGCAATCTACCAATGTTGTGACCTCGACCCCCAAGCCCGCGTGGCCATCAAG 8341

Qy       8089 TCGCTAACTGAGAGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGC 8148
Db       8342 TCCCTCACCGAGAGGCTTTATGTTGGGGGCCCTCTTACCAATTCAAGGGGGAGAACTGC 8401

Qy       8149 GGGTACAGGCGTTGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACA 8208
Db       8402 GGCTATCGCAGGTGCCGCGCGAGCGGCGTACTGACAACTAGCTGTGGTAACACCCTCACT 8461

Qy       8209 TGCTACGTGAAAGCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTG 8268
Db       8462 TGCTACATCAAGGCCCGGGCAGCCTGTCGAGCCGCAGGGCTCCAGGACTGCACCATGCTC 8521

Qy       8269 GTATGCGGCGATGACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGG 8328
Db       8522 GTGTGTGGCGACGACTTAGTCGTTATCTGTGAAAGCGCGGGGGTCCAGGAGGACGCGGCG 8581

Qy       8329 AACCTGAGAGCCTTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCC 8388
Db       8582 AGCCTGAGAGCCTTCACGGAGGCTATGACCAGGTACTCCGCCCCCCCTGGGGACCCCCCA 8641

Qy       8389 AGACCGGAGTATGATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTG 8448
Db       8642 CAACCAGAATACGACTTGGAGCTCATAACATCATGCTCCTCCAACGTGTCAGTCGCCCAC 8701

Qy       8449 GGCCCACAAGGCCGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGG 8508
Db       8702 GACGGCGCTGGAAAGAGGGTCTACTACCTCACCCGTGACCCTACAACCCCCCTCGCGAGA 8761

Qy       8509 GCTGCCTGGGAAACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAG 8568
Db       8762 GCTGCGTGGGAGACAGCAAGACACACTCCAGTCAATTCCTGGCTAGGCAACATAATCATG 8821

Qy       8569 TACGCCCCGACCATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATG 8628
Db       8822 TTTGCCCCCACACTGTGGGCGAGGATGATACTGATGACCCATTTCTTTAGCGTCCTTATA 8881

Qy       8629 GCTCAAGACACGCTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTG 8688
Db       8882 GCCAGGGACCAGCTTGAACAGGCCCTCGATTGCGAGATCTACGGGGCCTGCTACTCCATA 8941

Qy       8689 AGTCCCTTGGACCTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTG 8748
Db       8942 GAACCACTTGATCTACCTCCAATCATTCAAAGACTCCATGGCCTCAGCGCATTTTCACTC 9001

Qy       8749 CACACATACACTCCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCG 8808
Db       9002 CACAGTTACTCTCCAGGTGAAATTAATAGGGTGGCCGCATGCCTCAGAAAACTTGGGGTA 9061

Qy       8809 CCACCCCTCAGAGCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGT 8868
Db       9062 CCGCCCTTGCGAGCTTGGAGACACCGGGCCCGGAGCGTCCGCGCTAGGCTTCTGGCCAGA 9121

Qy       8869 GGGGGGAGAGCGGCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTC 8928
Db       9122 GGAGGCAGGGCTGCCATATGTGGCAAGTACCTCTTCAACTGGGCAGTAAGAACAAAGCTC 9181

Qy       8929 AAACTCACTCCATTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGC 8988
Db       9182 AAACTCACTCCAATAGCGGCCGCTGGCCAGCTGGACTTGTCCGGCTGGTTCACGGCTGGC 9241

Qy       8989 GCCGGCGGGGGCGACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTT 9048
Db       9242 TACAGCGGGGGAGACATTTATCACAGCGTGTCTCATGCCCGGCCCCGCTGGATCTGGTTT 9301

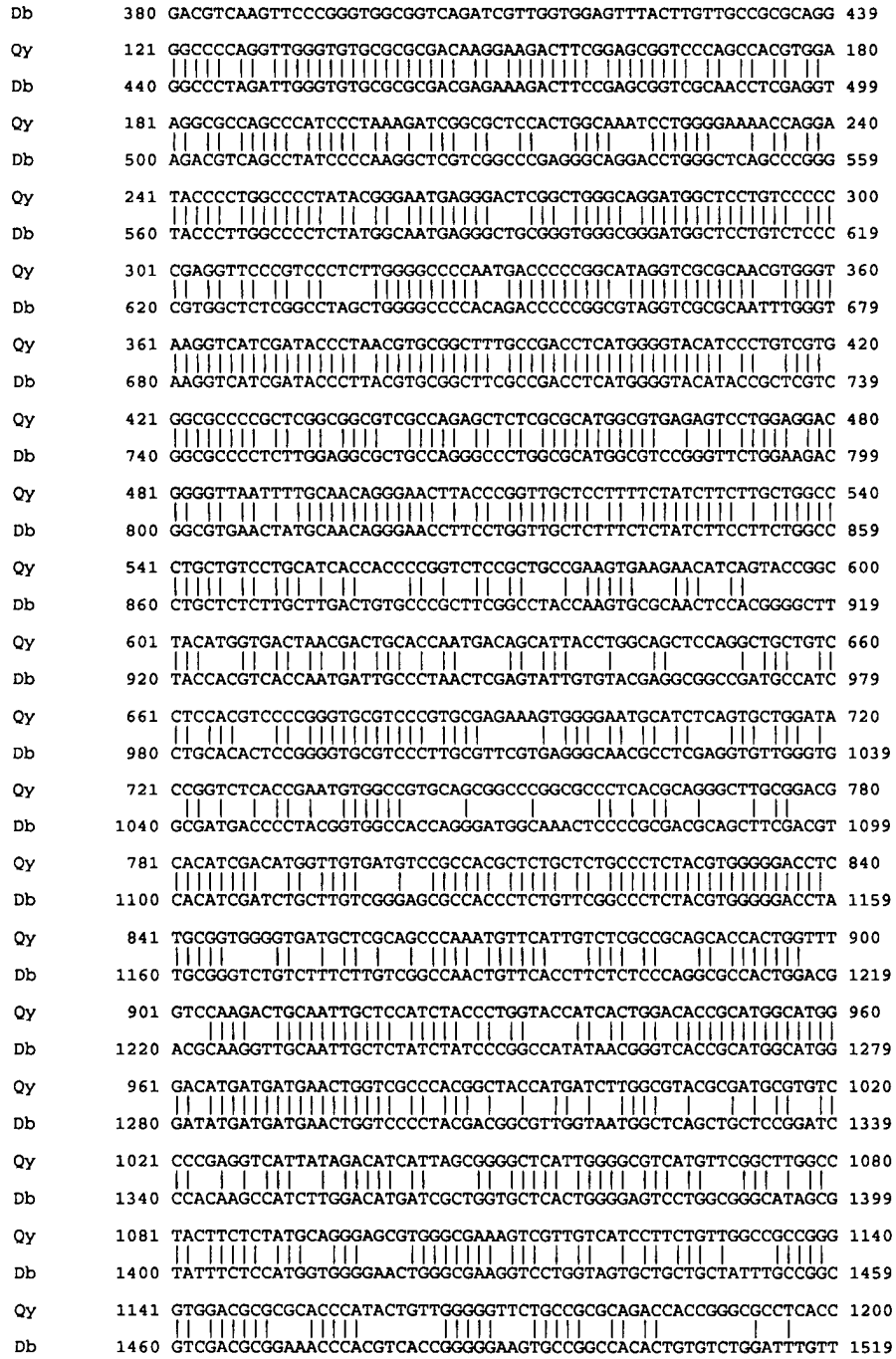
Qy       9049 GGCCTACTCCTACTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCG 9098
Db       9302 TGCCTACTCCTGCTTGCTGCAGGGGTAGGCATCTACCTCCTCCCCAACCG 9351
```

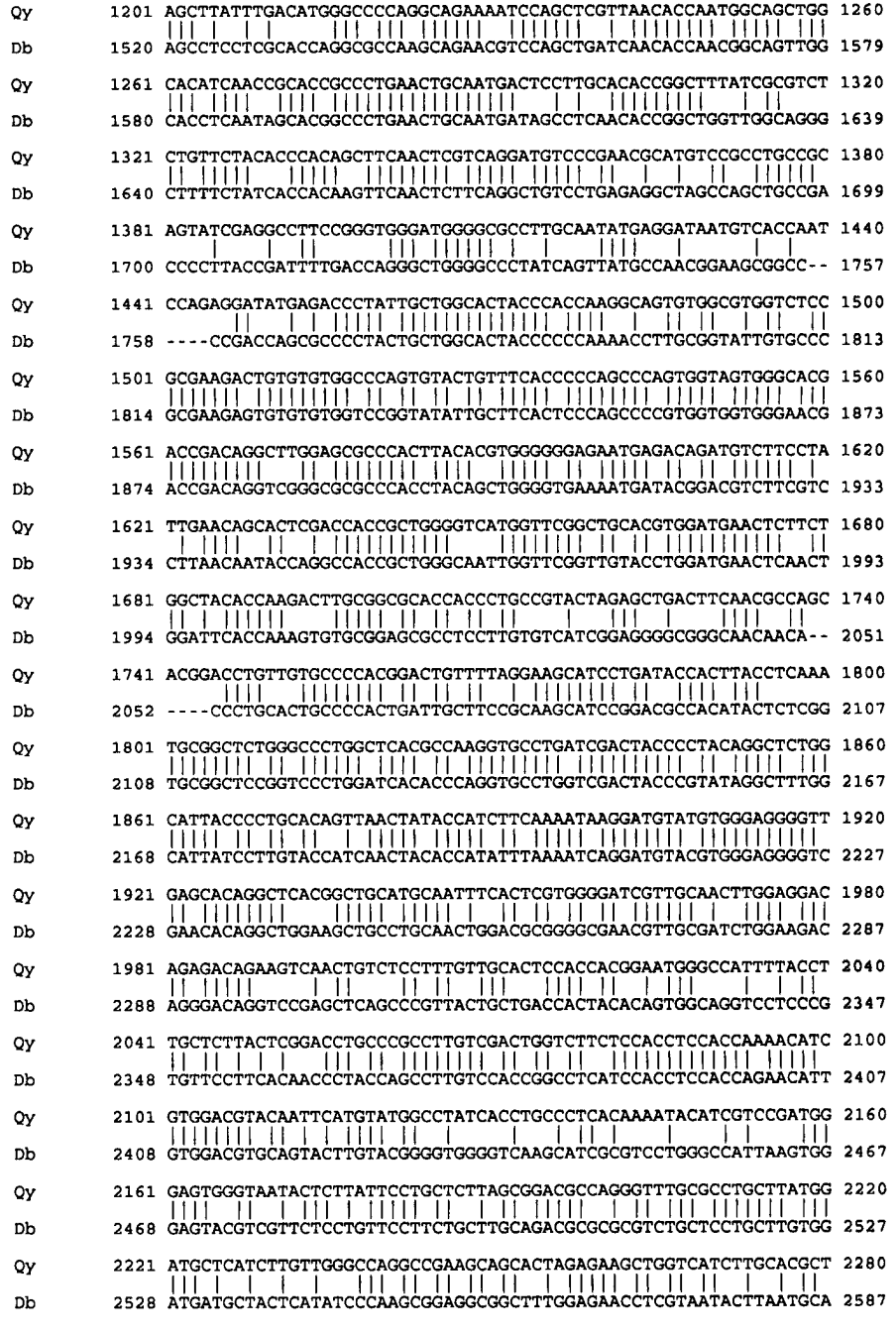
RESULT 14
US-09-916-359-1
; Sequence 1, Application US/09916359
; Patent No. 6538123
; GENERAL INFORMATION:
; APPLICANT: Veronique Barban
; TITLE OF INVENTION: VACCINE COMPOSITION FOR PREVENTING OR
; TITLE OF INVENTION: TREATING C HEPATITIS
; FILE REFERENCE: PMCF97-03A
; CURRENT APPLICATION NUMBER: US/09/916,359
; CURRENT FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: 09/388,874
; PRIOR FILING DATE: 1999-09-02
; PRIOR APPLICATION NUMBER: 97/02,887
; PRIOR FILING DATE: 1997-03-06
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 9379
; TYPE: DNA
; ORGANISM: Virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (320)...(9352)
US-09-916-359-1

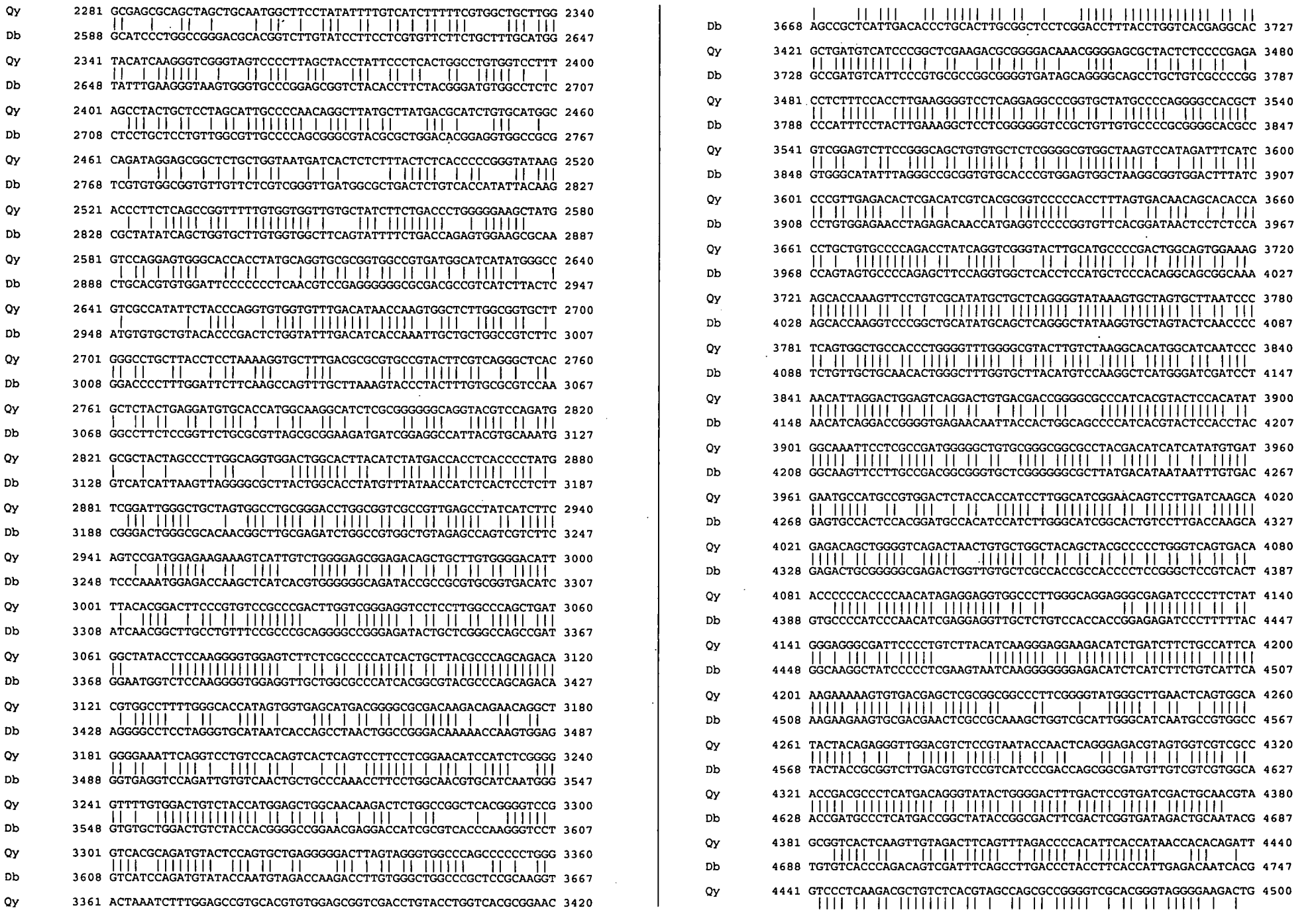
Query Match 45.2%; Score 4115.2; DB 4; Length 9379;
Best Local Similarity 66.3%; Pred. No. 0;
Matches 6042; Conservative 0; Mismatches 2978; Indels 90; Gaps 5;

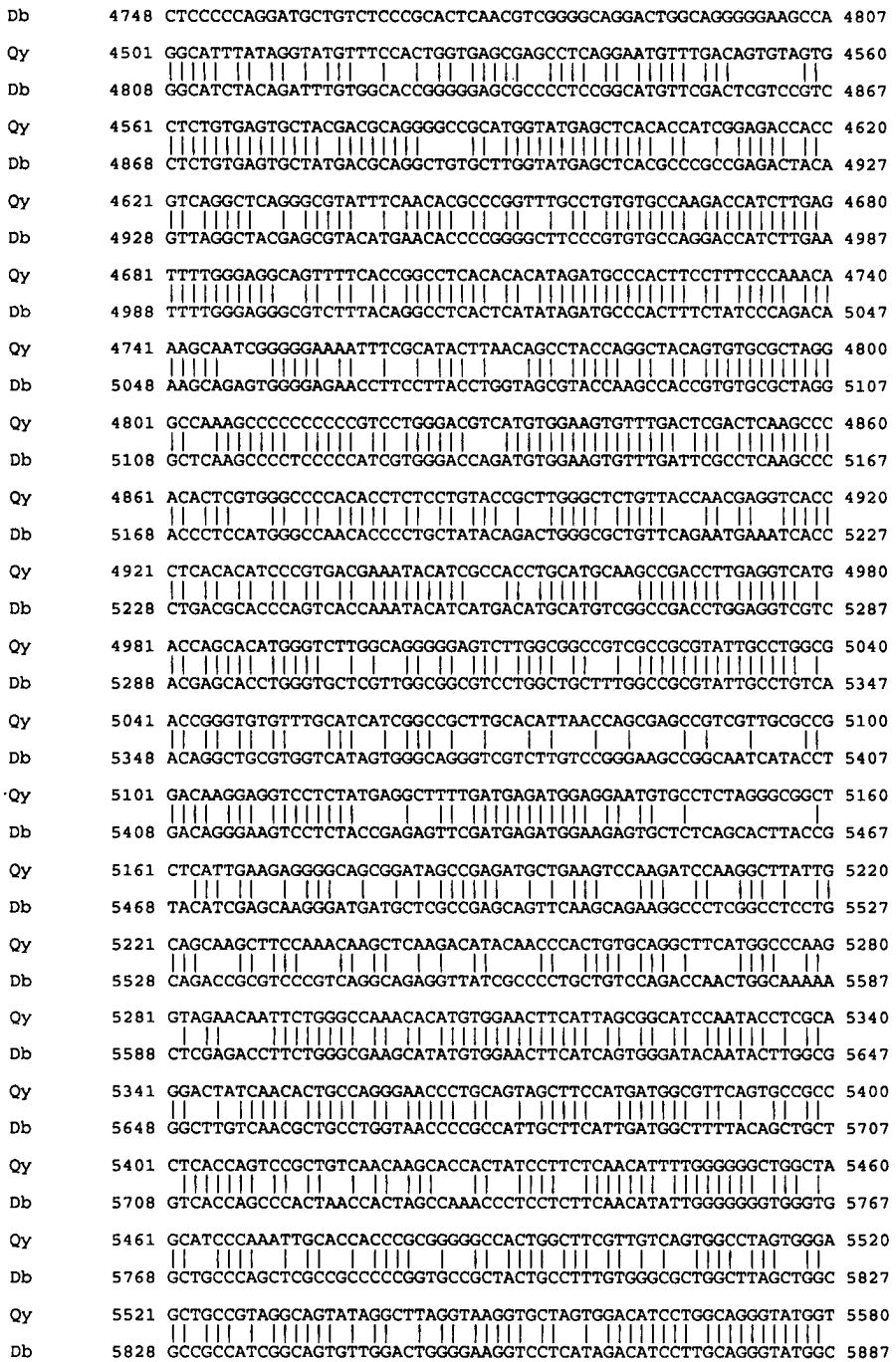
```
Qy          1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 60
Db        320 ATGAGCACGAATCCTAAACCTCAAAAAAAAAACAAACGTAACACCAACCGTCGCCCACAG 379

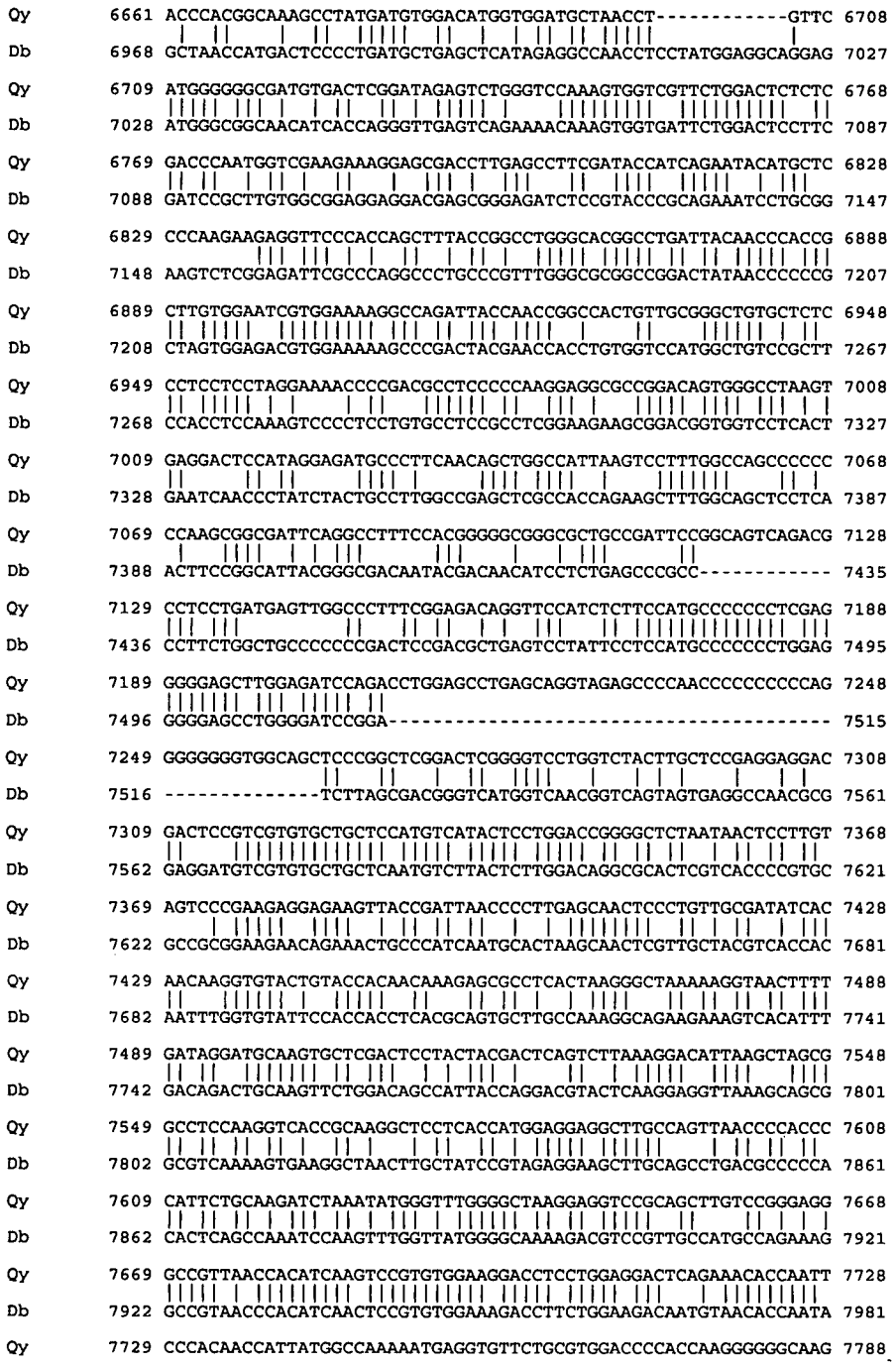
Qy         61 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 120
```


Db 380 GACGTCAAGTTCCCGGGTGGCGGTCAGATCGTTGGTGGAGTTTACTTGTTGCCGCGCAGG 439
Qy 121 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 180
Db 440 GGCCCTAGATTGGGTGTGCGCGCGACGAGAAAGACTTCCGAGCGGTCGCAACCTCGAGGT 499
Qy 181 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 240
Db 500 AGACGTCAGCCTATCCCCAAGGCTCGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGG 559
Qy 241 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 300
Db 560 TACCCTTGGCCCCTCTATGGCAATGAGGGCTGCGGGTGGGCGGGATGGCTCCTGTCTCCC 619
Qy 301 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 360
Db 620 CGTGGCTCTCGGCCTAGCTGGGGCCCCACAGACCCCCGGCGTAGGTCGCGCAATTTGGGT 679
Qy 361 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 420
Db 680 AAGGTCATCGATACCCTTACGTGCGGCTTCGCCGACCTCATGGGGTACATACCGCTCGTC 739
Qy 421 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 480
Db 740 GGCGCCCCTCTTGGAGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAAGAC 799
Qy 481 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 540
Db 800 GGCGTGAACTATGCAACAGGGAACCTTCCTGGTTGCTCTTTCTCTATCTTCCTTCTGGCC 859
Qy 541 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 600
Db 860 CTGCTCTCTTGCTTGACTGTGCCCGCTTCGGCCTACCAAGTGCGCAACTCCACGGGGCTT 919
Qy 601 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 660
Db 920 TACCACGTCACCAATGATTGCCCTAACTCGAGTATTGTGTACGAGGCGGCCGATGCCATC 979
Qy 661 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 720
Db 980 CTGCACACTCCGGGGGTGCGTCCCTTGCGTTCGTGAGGGCAACGCCTCGAGGTGTTGGGTG 1039
Qy 721 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 780
Db 1040 GCGATGACCCCTACGGTGGCCACCAGGGATGGCAAACTCCCCGCGACGCAGCTTCGACGT 1099
Qy 781 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 840
Db 1100 CACATCGATCTGCTTGTCGGGAGCGCCACCCTCTGTTCGGCCCTCTACGTGGGGGACCTA 1159
Qy 841 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 900
Db 1160 TGCGGGTCTGTCTTTCTTGTCGGCCAACTGTTCACCTTCTCTCCCAGGCGCCACTGGACG 1219
Qy 901 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 960
Db 1220 ACGCAAGGTTGCAATTGCTCTATCTATCCCGGCCATATAACGGGTCACCGCATGGCATGG 1279
Qy 961 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1020
Db 1280 GATATGATGATGAACTGGTCCCCTACGACGGCGTTGGTAATGGCTCAGCTGCTCCGGATC 1339
Qy 1021 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1080
Db 1340 CCACAAGCCATCTTGGACATGATCGCTGGTGCTCACTGGGGAGTCCTGGCGGGCATAGCG 1399
Qy 1081 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1140
Db 1400 TATTTCTCCATGGTGGGGAACTGGGCGAAGGTCCTGGTAGTGCTGCTGCTATTTGCCGGC 1459
Qy 1141 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1200
Db 1460 GTCGACGCGGAAACCCACGTCACCGGGGGAAGTGCCGGCCACACTGTGTCTGGATTTGTT 1519


Qy 1201 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1260
Db 1520 AGCCTCCTCGCACCAGGCGCCAAGCAGAACGTCCAGCTGATCAACACCAACGGCAGTTGG 1579
Qy 1261 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1320
Db 1580 CACCTCAATAGCACGGCCCTGAACTGCAATGATAGCCTCAACACCGGCTGGTTGGCAGGG 1639
Qy 1321 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1380
Db 1640 CTTTTCTATCACCACAAGTTCAACTCTTCAGGCTGTCCTGAGAGGCTAGCCAGCTGCCGA 1699
Qy 1381 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1440
Db 1700 CCCCTTACCGATTTTGACCAGGGCTGGGGCCCTATCAGTTATGCCAACGGAAGCGGCC-- 1757
Qy 1441 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1500
Db 1758 ----CCGACCAGCGCCCCTACTGCTGGCACTACCCCCCAAAACCTTGCGGTATTGTGCCC 1813
Qy 1501 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1560
Db 1814 GCGAAGAGTGTGTGTGGTCCGGTATATTGCTTCACTCCCAGCCCCGTGGTGGTGGGAACG 1873
Qy 1561 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1620
Db 1874 ACCGACAGGTCGGGCGCGCCCACCTACAGCTGGGGTGAAAATGATACGGACGTCTTCGTC 1933
Qy 1621 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 1680
Db 1934 CTTAACAATACCAGGCCACCGCTGGGCAATTGGTTCGGTTGTACCTGGATGAACTCAACT 1993
Qy 1681 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 1740
Db 1994 GGATTCACCAAAGTGTGCGGAGCGCCTCCTTGTGTCATCGGAGGGGCGGGCAACAACA-- 2051
Qy 1741 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 1800
Db 2052 ----CCCTGCACTGCCCCACTGATTGCTTCCGCAAGCATCCGGACGCCACATACTCTCGG 2107
Qy 1801 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 1860
Db 2108 TGCGGCTCCGGTCCCTGGATCACACCCAGGTGCCTGGTCGACTACCCGTATAGGCTTTGG 2167
Qy 1861 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 1920
Db 2168 CATTATCCTTGTACCATCAACTACACCATATTTAAAATCAGGATGTACGTGGGAGGGGTC 2227
Qy 1921 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 1980
Db 2228 GAACACAGGCTGGAAGCTGCCTGCAACTGGACGCGGGGCGAACGTTGCGATCTGGAAGAC 2287
Qy 1981 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2040
Db 2288 AGGGACAGGTCCGAGCTCAGCCCGTTACTGCTGACCACTACACAGTGGCAGGTCCTCCCG 2347
Qy 2041 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2100
Db 2348 TGTTCCTTCACAACCCTACCAGCCTTGTCCACCGGCCTCATCCACCTCCACCAGAACATT 2407
Qy 2101 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2160
Db 2408 GTGGACGTGCAGTACTTGTACGGGGTGGGGTCAAGCATCGCGTCCTGGGCCATTAAGTGG 2467
Qy 2161 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2220
Db 2468 GAGTACGTCGTTCTCCTGTTCCTTCTGCTTGCAGACGCGCGCGTCTGCTCCTGCTTGTGG 2527
Qy 2221 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2280
Db 2528 ATGATGCTACTCATATCCCAAGCGGAGGCGGCTTTGGAGAACCTCGTAATACTTAATGCA 2587




Db 4748 CTCCCCCAGGATGCTGTCTCCCGCACTCAACGTCGGGGCAGGACTGGCAGGGGGAAGCCA 4807
Qy 4501 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4560
Db 4808 GGCATCTACAGATTTGTGGCACCGGGGGAGCGCCCCTCCGGCATGTTCGACTCGTCCGTC 4867
Qy 4561 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4620
Db 4868 CTCTGTGAGTGCTATGACGCAGGCTGTGCTTGGTATGAGCTCACGCCCGCCGAGACTACA 4927
Qy 4621 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 4680
Db 4928 GTTAGGCTACGAGCGTACATGAACACCCCGGGGCTTCCCGTGTGCCAGGACCATCTTGAA 4987
Qy 4681 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 4740
Db 4988 TTTTGGGAGGGCGTCTTTACAGGCCTCACTCATATAGATGCCCACTTTCTATCCCAGACA 5047
Qy 4741 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 4800
Db 5048 AAGCAGAGTGGGGAGAACCTTCCTTACCTGGTAGCGTACCAAGCCACCGTGTGCGCTAGG 5107
Qy 4801 GCCAAAGCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 4860
Db 5108 GCTCAAGCCCCTCCCCCATCGTGGGACCAGATGTGGAAGTGTTTGATTCGCCTCAAGCCC 5167
Qy 4861 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 4920
Db 5168 ACCCTCCATGGGCCAACACCCCTGCTATACAGACTGGGCGCTGTTCAGAATGAAATCACC 5227
Qy 4921 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 4980
Db 5228 CTGACGCACCCAGTCACCAAATACATCATGACATGCATGTCGGCCGACCTGGAGGTCGTC 5287
Qy 4981 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5040
Db 5288 ACGAGCACCTGGGTGCTCGTTGGCGGCGTCCTGGCTGCTTTGGCCGCGTATTGCCTGTCA 5347
Qy 5041 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5100
Db 5348 ACAGGCTGCGTGGTCATAGTGGGCAGGGTCGTCTTGTCCGGGAAGCCGGCAATCATACCT 5407
Qy 5101 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5160
Db 5408 GACAGGGAAGTCCTCTACCGAGAGTTCGATGAGATGGAAGAGTGCTCTCAGCACTTACCG 5467
Qy 5161 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5220
Db 5468 TACATCGAGCAAGGGATGATGCTCGCCGAGCAGTTCAAGCAGAAGGCCCTCGGCCTCCTG 5527
Qy 5221 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5280
Db 5528 CAGACCGCGTCCCGTCAGGCAGAGGTTATCGCCCCTGCTGTCCAGACCAACTGGCAAAAA 5587
Qy 5281 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5340
Db 5588 CTCGAGACCTTCTGGGCGAAGCATATGTGGAACTTCATCAGTGGGATACAATACTTGGCG 5647
Qy 5341 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5400
Db 5648 GGCTTGTCAACGCTGCCTGGTAACCCCGCCATTGCTTCATTGATGGCTTTTACAGCTGCT 5707
Qy 5401 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5460
Db 5708 GTCACCAGCCCACTAACCACTAGCCAAACCCTCCTCTTCAACATATTGGGGGGGTGGGTG 5767
Qy 5461 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5520
Db 5768 GCTGCCCAGCTCGCCGCCCCCGGTGCCGCTACTGCCTTTGTGGGCGCTGGCTTAGCTGGC 5827
Qy 5521 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5580
Db 5828 GCCGCCATCGGCAGTGTTGGACTGGGGAAGGTCCTCATAGACATCCTTGCAGGGTATGGC 5887
Qy 5581 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5640
Db 5888 GCGGGCGTGGCGGGAGCTCTTGTGGCATTCAAGATCATGAGCGGTGAGGTCCCCTCCACG 5947
Qy 5641 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 5700
Db 5948 GAGGACCTGGTCAATCTACTGCCCGCCATCCTCTCGCCCGGAGCCCTCGTAGTCGGCGTG 6007
Qy 5701 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 5760
Db 6008 GTCTGTGCAGCAATACTGCGCCGGCACGTTGGCCCGGGCGAGGGGGCAGTGCAGTGGATG 6067
Qy 5761 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 5820
Db 6068 AACCGGCTGATAGCCTTCGCCTCCCGGGGGAACCATGTTTCCCCCACGCACTACGTGCCG 6127
Qy 5821 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 5880
Db 6128 GAGAGCGATGCAGCTGCCCGCGTCACTGCCATACTCAGCAGCCTCACTGTAACCCAGCTC 6187
Qy 5881 CTCAGAAGACTCCACAACTGGATTACTGAGGACTGCCCCATCCCATGCGGCGGCTCGTGG 5940
Db 6188 CTGAGGCGACTGCACCAGTGGATAAGCTCGGAGTGTACCACTCCATGCTCCGGTTCCTGG 6247
Qy 5941 CTCCGCGATGTGTGGGACTGGGTTTGCACCATCCTAACAGACTTTAAAAATTGGCTGACC 6000
Db 6248 CTAAGGGACATCTGGGACTGGATATGCGAGGTGTTGAGCGACTTTAAGACCTGGCTAAAA 6307
Qy 6001 TCCAAATTATTCCCAAAGATGCCCGGCCTCCCCTTTGTCTCCTGTCAAAAGGGGTACAAG 6060
Db 6308 GCTAAGCTCATGCCACAGCTGCCTGGGATCCCCTTTGTGTCCTGCCAGCGCGGGTATAAG 6367
Qy 6061 GGCGTGTGGGCCGGCACTGGCATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6120
Db 6368 GGGGTCTGGCGAGTGGACGGCATCATGCACACTCGCTGCCACTGTGGAGCTGAGATCACT 6427
Qy 6121 GGCAATGTCCGCTTGGGCTCCATGAGAATCACGGGGCCTAAGACCTGCATGAATATCTGG 6180
Db 6428 GGACATGTCAAAAACGGGACGATGAGGATCGTCGGTCCTAGGACCTGCAGGAACATGTGG 6487
Qy 6181 CAGGGGACCTTTCCTATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCGCCA 6240
Db 6488 AGTGGGACCTTCCCCATTAATGCCTACACCACGGGCCCCTGTACCCCCCTTCCTGCGCCG 6547
Qy 6241 AACTTTAAGGTCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6300
Db 6548 AACTACACGTTCGCGCTATGGAGGGTGTCTGCAGAGGAATATGTGGAGATAAGGCAGGTG 6607
Qy 6301 GGGTCATACCACTACATAACAGGACTCACCACTGATAACTTGAAAGTCCCCTGCCAACTA 6360
Db 6608 GGGGACTTCCACTACGTGACGGGTATGACTACTGACAATCTCAAATGCCCGTGCCAGGTC 6667
Qy 6361 CCCTCTCCCGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCACA 6420
Db 6668 CCATCGCCCGAATTTTTCACAGAATTGGACGGGGTGCGCCTACATAGGTTTGCGCCCCCC 6727
Qy 6421 CCGAAGCCGTTTTTCCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6480
Db 6728 TGCAAGCCCTTGCTGCGGGAGGAGGTATCATTCAGAGTAGGACTCCACGAATACCCGGTA 6787
Qy 6481 GGGTCCCAGCTTCCTTGCGACCCTGAACCCGACACAGACGTATTGATGTCCATGCTAACA 6540
Db 6788 GGGTCGCAATTACCTTGCGAGCCCGAACCGGACGTGGCCGTGTTGACGTCCATGCTCACT 6847
Qy 6541 GATCCATCTCATATCACGGCGGAGACTGCAGCGCGGCGTTTAGCGCGGGGGTCACCCCCA 6600
Db 6848 GATCCCTCCCATATAACAGCAGAGGCGGCCGGGCGAAGGTTGGCGAGGGGATCACCCCCC 6907
Qy 6601 TCCGAGGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 6660
Db 6908 TCTGTGGCCAGCTCCTCGGCTAGCCAGCTATCCGCTCCATCTCTCAAGGCAACTTGCACC 6967


Qy 6661 ACCCACGGCAAAGCCTATGATGTGGACATGGTGGATGCTAACCT----------GTTC 6708
Db 6968 GCTAACCATGACTCCCCTGATGCTGAGCTCATAGAGGCCAACCTCCTATGGAGGCAGGAG 7027
Qy 6709 ATGGGGGGCGATGTGACTCGGATAGAGTCTGGGTCCAAAGTGGTCGTTCTGGACTCTCTC 6768
Db 7028 ATGGGCGGCAACATCACCAGGGTTGAGTCAGAAAACAAAGTGGTGATTCTGGACTCCTTC 7087
Qy 6769 GACCCAATGGTCGAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCAGAATACATGCTC 6828
Db 7088 GATCCGCTTGTGGCGGAGGAGGACGAGCGGGAGATCTCCGTACCCGCAGAAATCCTGCGG 7147
Qy 6829 CCCAAGAAGAGGTTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCG 6888
Db 7148 AAGTCTCGGAGATTCGCCCAGGCCCTGCCCGTTTGGGCGCGGCCGGACTATAACCCCCCG 7207
Qy 6889 CTTGTGGAATCGTGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTC 6948
Db 7208 CTAGTGGAGACGTGGAAAAAGCCCGACTACGAACCACCTGTGGTCCATGGCTGTCCGCTT 7267
Qy 6949 CCTCCTCCTAGGAAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGT 7008
Db 7268 CCACCTCCAAAGTCCCCTCCTGTGCCTCCGCCTCGGAAGAAGCGGACGGTGGTCCTCACT 7327
Qy 7009 GAGGACTCCATAGGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCC 7068
Db 7328 GAATCAACCCTATCTACTGCCTTGGCCGAGCTCGCCACCAGAAGCTTTGGCAGCTCCTCA 7387
Qy 7069 CCAAGCGGCGATTCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACG 7128
Db 7388 ACTTCCGGCATTACGGGCGACAATACGACAACATCCTCTGAGCCCGCC---------- 7435
Qy 7129 CCTCCTGATGAGTTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCTCGAG 7188
Db 7436 CCTTCTGGCTGCCCCCCCGACTCCGACGCTGAGTCCTATTCCTCCATGCCCCCCCTGGAG 7495
Qy 7189 GGGGAGCTTGGAGATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCAG 7248
Db 7496 GGGGAGCCTGGGGATCCGGA---------- 7515
Qy 7249 GGGGGGGTGGCAGCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGAC 7308
Db 7516 ----------TCTTAGCGACGGGTCATGGTCAACGGTCAGTAGTGAGGCCAACGCG 7561
Qy 7309 GACTCCGTCGTGTGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGT 7368
Db 7562 GAGGATGTCGTGTGCTGCTCAATGTCTTACTCTTGGACAGGCGCACTCGTCACCCCGTGC 7621
Qy 7369 AGTCCCGAAGAGGAGAAGTTACCGATTAACCCCTTGAGCAACTCCCTGTTGCGATATCAC 7428
Db 7622 GCCGCGGAAGAACAGAAACTGCCCATCAATGCACTAAGCAACTCGTTGCTACGTCACCAC 7681
Qy 7429 AACAAGGTGTACTGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTT 7488
Db 7682 AATTTGGTGTATTCCACCACCTCACGCAGTGCTTGCCAAAGGCAGAAGAAAGTCACATTT 7741
Qy 7489 GATAGGATGCAAGTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCG 7548
Db 7742 GACAGACTGCAAGTTCTGGACAGCCATTACCAGGACGTACTCAAGGAGGTTAAAGCAGCG 7801
Qy 7549 GCCTCCAAGGTCACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCC 7608
Db 7802 GCGTCAAAAGTGAAGGCTAACTTGCTATCCGTAGAGGAAGCTTGCAGCCTGACGCCCCCA 7861
Qy 7609 CATTCTGCAAGATCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGG 7668
Db 7862 CACTCAGCCAAATCCAAGTTTGGTTATGGGGCAAAAGACGTCCGTTGCCATGCCAGAAAG 7921
Qy 7669 GCCGTTAACCACATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATT 7728
Db 7922 GCCGTAACCCACATCAACTCCGTGTGGAAAGACCTTCTGGAAGACAATGTAACACCAATA 7981
Qy 7729 CCCACAACCATTATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAG 7788
Db 7982 GACACTACCATCATGGCTAAGAACGAGGTTTTCTGCGTTCAGCCTGAGAAGGGGGGTCGT 8041
Qy 7789 AAAGCAGCTCGCCTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCC 7848
Db 8042 AAGCCAGCTCGTCTCATCGTGTTCCCCGATCTGGGCGTGCGCGTGTGCGAAAAGATGGCT 8101
Qy 7849 CTTTATGACATTACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAG 7908
Db 8102 TTGTACGACGTGGTTACAAAGCTCCCCTTGGCCGTGATGGGAAGCTCCTACGGATTCCAA 8161
Qy 7909 TATTCCCCCGCTCAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCT 7968
Db 8162 TACTCACCAGGACAGCGGGTTGAATTCCTCGTGCAAGCGTGGAAGTCCAAGAAAACCCCA 8221
Qy 7969 ATGGGTTTTTCGTATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGG 8028
Db 8222 ATGGGGTTCTCGTATGATACCCGCTGCTTTGACTCCACAGTCACTGAGAGCGACATCCGT 8281
Qy 8029 ACTGAGGAGTCCATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACAC 8088
Db 8282 ACGGAGGAGGCAATCTACCAATGTTGTGACCTCGACCCCCAAGCCCGCGTGGCCATCAAG 8341
Qy 8089 TCGCTAACTGAGAGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGC 8148
Db 8342 TCCCTCACCGAGAGGCTTTATGTTGGGGGCCCTCTTACCAATTCAAGGGGGGAGAACTGC 8401
Qy 8149 GGGTACAGGCGTTGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACA 8208
Db 8402 GGCTATCGCAGGTGCCGCGCGAGCGGCGTACTGACAACTAGCTGTGGTAACACCCTCACT 8461
Qy 8209 TGCTACGTGAAAGCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTG 8268
Db 8462 TGCTACATCAAGGCCCGGGCAGCCTGTCGAGCCGCAGGGCTCCAGGACTGCACCATGCTC 8521
Qy 8269 GTATGCGGCGATGACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGG 8328
Db 8522 GTGTGTGGCGACGACTTAGTCGTTATCTGTGAAAGCGCGGGGGTCCAGGAGGACGCGGCG 8581
Qy 8329 AACCTGAGAGCCTTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCC 8388
Db 8582 AGCCTGAGAGCCTTCACGGAGGCTATGACCAGGTACTCCGCCCCCCCCTGGGGACCCCCCA 8641
Qy 8389 AGACCGGAGTATGATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTG 8448
Db 8642 CAACCAGAATACGACTTGGAGCTCATAACATCATGCTCCTCCAACGTGTCAGTCGCCCAC 8701
Qy 8449 GGCCCACAAGGCCGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGG 8508
Db 8702 GACGGCGCTGGAAAGAGGGTCTACTACCTCACCCGTGACCCTACAACCCCCCTCGCGAGA 8761
Qy 8509 GCTGCCTGGGAAACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAG 8568
Db 8762 GCTGCGTGGGAGACAGCAAGACACACTCCAGTCAATTCCTGGCTAGGCAACATAATCATG 8821
Qy 8569 TACGCCCCGACCATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATG 8628
Db 8822 TTTGCCCCCACACTGTGGGCGAGGATGATACTGATGACCCATTTCTTTAGCGTCCTTATA 8881
Qy 8629 GCTCAAGACACGCTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTG 8688
Db 8882 GCCAGGGACCAGCTTGAACAGGCCCTCGATTGCGAGATCTACGGGGCCTGCTACTCCATA 8941
Qy 8689 AGTCCCTTGGACCTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTG 8748
Db 8942 GAACCACTTGATCTACCTCCAATCATTCAAAGACTCCATGGCCTCAGCGCATTTTCACTC 9001
Qy 8749 CACACATACACTCCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCG 8808
Db 9002 CACAGTTACTCTCCAGGTGAAATTAATAGGGTGGCCGCATGCCTCAGAAAACTTGGGGTA 9061
Qy 8809 CCACCCCTCAGAGCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGT 8868

```
Db       9062 CCGCCCTTGCGAGCTTGGAGACACCGGGCCCGGAGCGTCCGCGCTAGGCTTCTGGCCAGA 9121

Qy       8869 GGGGGGAGAGCGGCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTC 8928
              || || || || ||| | || ||     || |||||||| ||||| || |  || ||||||
Db       9122 GGAGGCAGGGCTGCCATATGTGGCAAGTACCTCTTCAACTGGGCAGTAAGAACAAAGCTC 9181

Qy       8929 AAACTCACTCCATTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGC 8988
              ||||||||||||  |  |||  ||  |||  ||||| || ||| ||||||||||  |  |||
Db       9182 AAACTCACTCCAATAGCGGCCGCTGGCCAGCTGGACTTGTCCGGCTGGTTCACGGCTGGC 9241

Qy       8989 GCCGGCGGGGGCGACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTT 9048
                | ||||||| ||||||||||||||||||||| | ||||||  |||||||   |    |||
Db       9242 TACAGCGGGGGAGACATTTATCACAGCGTGTCTCATGCCCGGCCCCGCTGGATCTGGTTT 9301

Qy       9049 GGCCTACTCCTACTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCG 9098
              ||||||||||  |||  ||  |||||||||| | | |||  ||||||   ||
Db       9302 TGCCTACTCCTGCTTGCTGCAGGGGTAGGCATCTACCTCCTCCCCAACCG 9351


RESULT 15
US-07-910-760-9
; Sequence 9, Application US/07910760
; Patent No. 5683864
;  GENERAL INFORMATION:
;    APPLICANT:  Houghton, Michael
;    APPLICANT:  Choo, Qui-Lim
;    APPLICANT:  Kuo, George
;    TITLE OF INVENTION:  Combinations of Hepatitis C virus (HCV)
;    TITLE OF INVENTION:  Antigens for Use in Immunoassays for Anti-HCV Antibodies
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Chiron Corporation
;      STREET:  P.O. Box 8097  (Int. Prop. R-440)
;      CITY:  Emeryville
;      STATE:  CA
;      COUNTRY:  U.S.A.
;      ZIP:  94662-8097
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/910,760
;      FILING DATE:  07-JUL-1992
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Blackburn Esq., Robert P.
;      REGISTRATION NUMBER:  30,447
;      REFERENCE/DOCKET NUMBER:  0101.002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (510) 601-2702
;      TELEFAX:  (510) 655-3542
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9401 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  342..9374
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  366
;      OTHER INFORMATION:  /note= "This amino acid position
;      OTHER INFORMATION:  can also be Arg."
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  372
;      OTHER INFORMATION:  /note= "This amino acid position
;      OTHER INFORMATION:  can also be Thr."
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  867
;      OTHER INFORMATION:  /note= "This amino acid position
;      OTHER INFORMATION:  can also be Thr."
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1341
;      OTHER INFORMATION:  /note= "This amino acid position
;      OTHER INFORMATION:  can also be Val."
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  2148
;      OTHER INFORMATION:  /note= "This amino acid position
;      OTHER INFORMATION:  can also be Ile."
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  2883
;      OTHER INFORMATION:  /note= "This amino acid position
;      OTHER INFORMATION:  can also be Asn."
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  3681
;      OTHER INFORMATION:  /note= "This amino acid position
;      OTHER INFORMATION:  can also be Ser."
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  3690
;      OTHER INFORMATION:  /note= "This amino acid position
;      OTHER INFORMATION:  can also be Thr."
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  4167
;      OTHER INFORMATION:  /note= "This amino acid position
;      OTHER INFORMATION:  can also be Leu."
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  4323
;      OTHER INFORMATION:  /note= "This amino acid position
;      OTHER INFORMATION:  can also be Val."
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  4701
;      OTHER INFORMATION:  /note= "This amino acid position
;      OTHER INFORMATION:  can also be Tyr."
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  4752
;      OTHER INFORMATION:  /note= "This amino acid position
;      OTHER INFORMATION:  can also be Ser."
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  5970
;      OTHER INFORMATION:  /note= "This amino acid position
;      OTHER INFORMATION:  can also be Gly."
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6183
;      OTHER INFORMATION:  /note= "This amino acid position
;      OTHER INFORMATION:  can also be His."
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6186
;      OTHER INFORMATION:  /note= "This amino acid position
;      OTHER INFORMATION:  can also be Cys."
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6402
;      OTHER INFORMATION:  /note= "This amino acid position
;      OTHER INFORMATION:  can also be Val."
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  7386
```

```
;       OTHER INFORMATION:  /note= "This amino acid position
;       OTHER INFORMATION:  can also be Ser."
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  7494
;       OTHER INFORMATION:  /note= "This amino acid position
;       OTHER INFORMATION:  can also be Phe."
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  7497
;       OTHER INFORMATION:  /note= "This amino acid position
;       OTHER INFORMATION:  can also be Ala."
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  7845
;       OTHER INFORMATION:  /note= "This amino acid position
;       OTHER INFORMATION:  can also be Phe."
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  8409
;       OTHER INFORMATION:  /note= "This amino acid position
;       OTHER INFORMATION:  can also be Gly."
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  9102
;       OTHER INFORMATION:  /note= "This amino acid position
;       OTHER INFORMATION:  can also by Gly."
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  9327
;       OTHER INFORMATION:  /note= "This amino acid position
;       OTHER INFORMATION:  can also be Pro."
US-07-910-760-9

  Query Match             45.2%;  Score 4115.2;  DB 1;  Length 9401;
  Best Local Similarity   66.3%;  Pred. No. 0;
  Matches 6042;  Conservative    0;  Mismatches 2978;  Indels   90;  Gaps    5;

Qy          1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGTCGCCCACAA 60
              |||||||| |||||||||||||||| ||||| |||| | ||||||||||||||||||||
Db        342 ATGAGCACGAATCCTAAACCTCAAAAAAAAAACAAACGTAACACCAACCGTCGCCCACAG 401

Qy         61 GACGTTAAGTTTCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGG 120
              ||||| ||||| ||||| ||||| |||||||||| ||||| ||||||||||||||||||
Db        402 GACGTCAAGTTCCCGGGTGGCGGTCAGATCGTTGGTGGAGTTTACTTGTTGCCGCGCAGG 461

Qy        121 GGCCCCAGGTTGGGTGTGCGCGCGACAAGGAAGACTTCGGAGCGGTCCCAGCCACGTGGA 180
              ||||| || |||||||||||||||| || |||||||| |||||||| || || || ||
Db        462 GGCCCTAGATTGGGTGTGCGCGCGACGAGAAAGACTTCCGAGCGGTCGCAACCTCGAGGT 521

Qy        181 AGGCGCCAGCCCATCCCTAAAGATCGGCGCTCCACTGGCAAATCCTGGGGAAAACCAGGA 240
              || || ||||| ||||| || | ||| ||  ||   ||||   ||||||   | || ||
Db        522 AGACGTCAGCCTATCCCCAAGGCTCGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGG 581

Qy        241 TACCCCTGGCCCCTATACGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCC 300
              ||||| |||||||| || || ||||||||    ||| ||||| |||||||||||||| |||
Db        582 TACCCTTGGCCCCTCTATGGCAATGAGGGCTGCGGGTGGGCGGGATGGCTCCTGTCTCCC 641

Qy        301 CGAGGTTCCCGTCCCTCTTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAACGTGGGT 360
              || || || || ||    |||||||||  |||||||||  ||||||||||||   |||||
Db        642 CGTGGCTCTCGGCCTAGCTGGGGCCCCACAGACCCCCGGCGTAGGTCGCGCAATTTGGGT 701

Qy        361 AAGGTCATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCTGTCGTG 420
              |||||||||||||||| |||||||||| |||||||||||||||||||| ||   ||||
Db        702 AAGGTCATCGATACCCTTACGTGCGGCTTCGCCGACCTCATGGGGTACATACCGCTCGTC 761

Qy        421 GGCGCCCCGCTCGGCGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGAC 480
              |||||||| || || ||||   ||||| || || ||||||||||   | || |||||  |||
Db        762 GGCGCCCCTCTTGGAGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAAGAC 821

Qy        481 GGGGTTAATTTTGCAACAGGGAACTTACCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCC 540
              || || || | |||||||||||||| | || |||||||| || |||||||| | ||||||
Db        822 GGCGTGAACTATGCAACAGGGAACCTTCCTGGTTGCTCTTTCTCTATCTTCCTTCTGGCC 881

Qy        541 CTGCTGTCCTGCATCACCACCCCGGTCTCCGCTGCCGAAGTGAAGAACATCAGTACCGGC 600
              ||||| || ||| | ||    || |  || ||   | |||||   |||  ||
Db        882 CTGCTCTCTTGCTTGACTGTGCCCGCTTCGGCCTACCAAGTGCGCAACTCCACGGGGCTT 941

Qy        601 TACATGGTGACTAACGACTGCACCAATGACAGCATTACCTGGCAGCTCCAGGCTGCTGTC 660
              ||| || || || || ||| | ||    || |||   |   ||      | |||  ||
Db        942 TACCACGTCACCAATGATTGCCCTAACTCGAGTATTGTGTACGAGGCGGCCGATGCCATC 1001

Qy        661 CTCCACGTCCCCGGGTGCGTCCCGTGCGAGAAAGTGGGGAATGCATCTCAGTGCTGGATA 720
              || |||   || |||||||||||| ||||      | ||| || || ||   ||| ||| |
Db       1002 CTGCACACTCCGGGGTGCGTCCCTTGCGTTCGTGAGGGCAACGCCTCGAGGTGTTGGGTG 1061

Qy        721 CCGGTCTCACCGAATGTGGCCGTGCAGCGGCCCGGCGCCCTCACGCAGGGCTTGCGGACG 780
              || | | || |  ||||||      |      |     || | ||  |    | ||
Db       1062 GCGATGACCCCTACGGTGGCCACCAGGGATGGCAAACTCCCCGCGACGCAGCTTCGACGT 1121

Qy        781 CACATCGACATGGTTGTGATGTCCGCCACGCTCTGCTCTGCCCTCTACGTGGGGGACCTC 840
              |||||||  || ||||  | |||||| |||| || ||||||||||||||||||||
Db       1122 CACATCGATCTGCTTGTCGGGAGCGCCACCCTCTGTTCGGCCCTCTACGTGGGGGACCTA 1181

Qy        841 TGCGGTGGGGTGATGCTCGCAGCCCAAATGTTCATTGTCTCGCCGCAGCACCACTGGTTT 900
              |||||    ||  | || |  | |||| ||||||   |||| ||   || |||||||
Db       1182 TGCGGGTCTGTCTTTCTTGTCGGCCAACTGTTCACCTTCTCTCCCAGGCGCCACTGGACG 1241

Qy        901 GTCCAAGACTGCAATTGCTCCATCTACCCTGGTACCATCACTGGACACCGCATGGCATGG 960
                |||| |||||||||| ||||| || ||    || || || ||||||||||||||||
Db       1242 ACGCAAGGTTGCAATTGCTCTATCTATCCCGGCCATATAACGGGTCACCGCATGGCATGG 1301

Qy        961 GACATGATGATGAACTGGTCGCCCACGGCTACCATGATCTTGGCGTACGCGATGCGTGTC 1020
              || |||||||||||||||| || ||| |  |  || |  ||||  |   | | ||  ||
Db       1302 GATATGATGATGAACTGGTCCCCTACGACGGCGTTGGTAATGGCTCAGCTGCTCCGGATC 1361

Qy       1021 CCCGAGGTCATTATAGACATCATTAGCGGGGCTCATTGGGGCGTCATGTTCGGCTTGGCC 1080
              ||  | | |||  | ||||| ||    || ||||| |||| ||| ||   ||| | ||
Db       1362 CCACAAGCCATCTTGGACATGATCGCTGGTGCTCACTGGGGAGTCCTGGCGGGCATAGCG 1421

Qy       1081 TACTTCTCTATGCAGGGAGCGTGGGCGAAAGTCGTTGTCATCCTTCTGTTGGCCGCCGGG 1140
              || ||||| |||  |||    |||||||| ||| | ||  | || ||| |    |||||
Db       1422 TATTTCTCCATGGTGGGGAACTGGGCGAAGGTCCTGGTAGTGCTGCTGCTATTTGCCGGC 1481

Qy       1141 GTGGACGCGCGCACCCATACTGTTGGGGGTTCTGCCGCGCAGACCACCGGGCGCCTCACC 1200
              || ||||||   |||||        |||||  |||||  || ||       | |
Db       1482 GTCGACGCGGAAACCCACGTCACCGGGGGAAGTGCCGGCCACACTGTGTCTGGATTTGTT 1541

Qy       1201 AGCTTATTTGACATGGGCCCCAGGCAGAAAATCCAGCTCGTTAACACCAATGGCAGCTGG 1260
              ||| |  | |     ||| ||| ||||||  |||||||  | |||||||| ||||| |||
Db       1542 AGCCTCCTCGCACCAGGCGCCAAGCAGAACGTCCAGCTGATCAACACCAACGGCAGTTGG 1601

Qy       1261 CACATCAACCGCACCGCCCTGAACTGCAATGACTCCTTGCACACCGGCTTTATCGCGTCT 1320
              ||| ||||  |||| |||||||||||||||||    | |  |||||||||   | ||
Db       1602 CACCTCAATAGCACGGCCCTGAACTGCAATGATAGCCTCAACACCGGCTGGTTGGCAGGG 1661

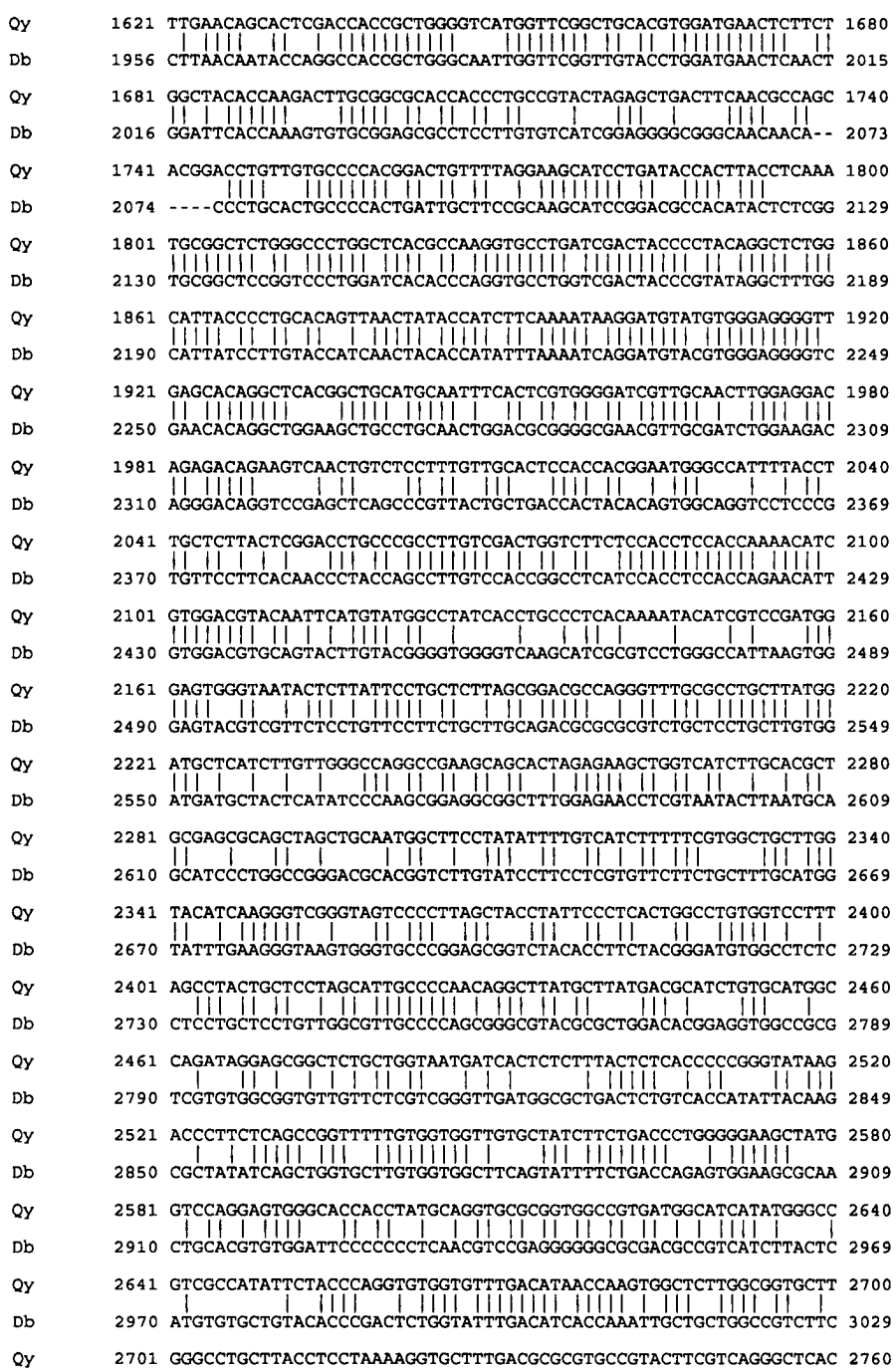
Qy       1321 CTGTTCTACACCCACAGCTTCAACTCGTCAGGATGTCCCGAACGCATGTCCGCCTGCCGC 1380
              || |||||   |||||  |||||||| ||||| ||||| ||  |  |  ||  ||||||
Db       1662 CTTTTCTATCACCACAAGTTCAACTCTTCAGGCTGTCCTGAGAGGCTAGCCAGCTGCCGA 1721

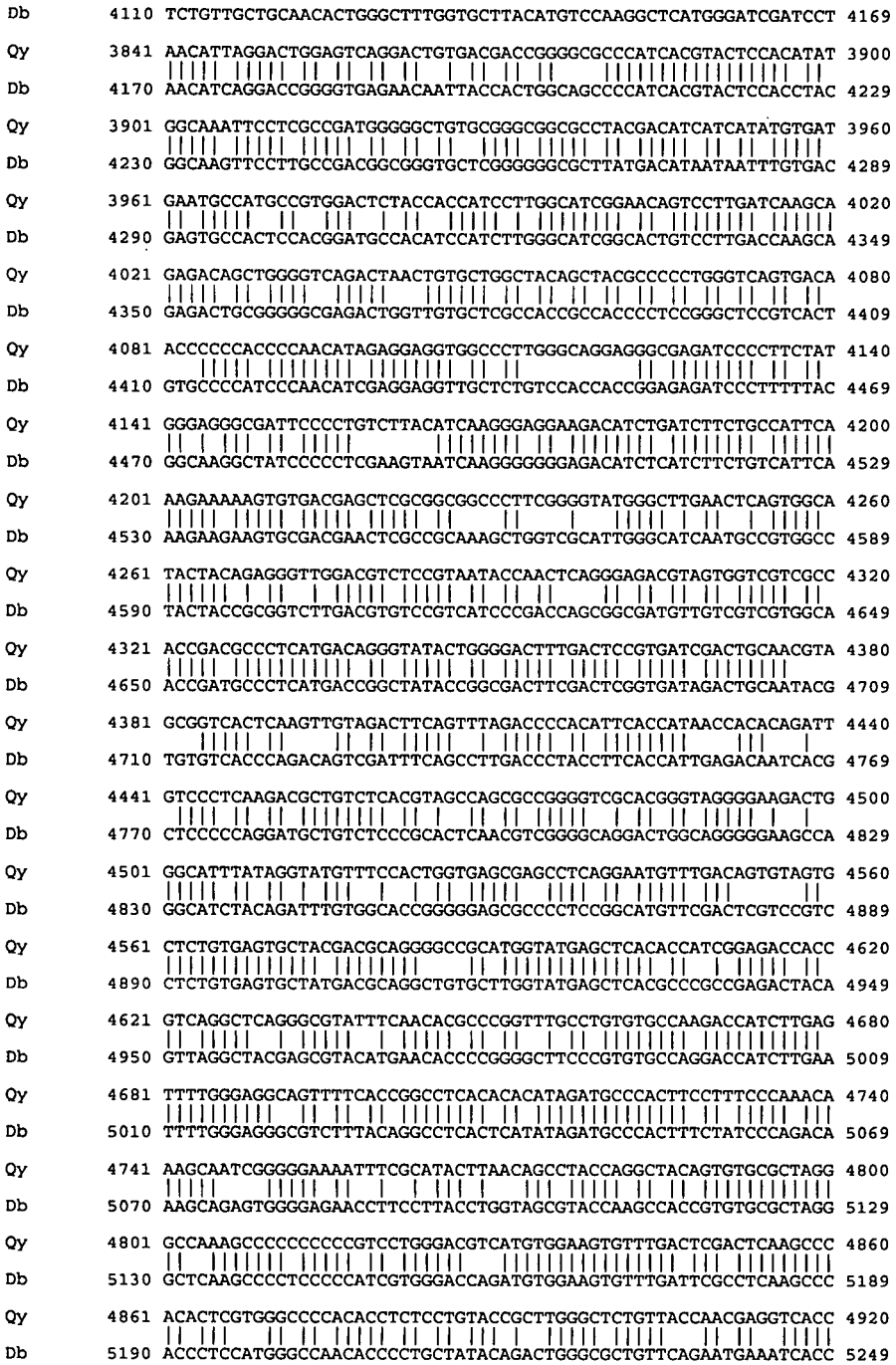
Qy       1381 AGTATCGAGGCCTTCCGGGTGGGATGGGGCGCCTTGCAATATGAGGATAATGTCACCAAT 1440
                |    |  ||    ||| |||||| |  |    ||||   |      |  |
Db       1722 CCCCTTACCGATTTTGACCAGGGCTGGGGCCCTATCAGTTATGCCAACGGAAGCGGCC-- 1779

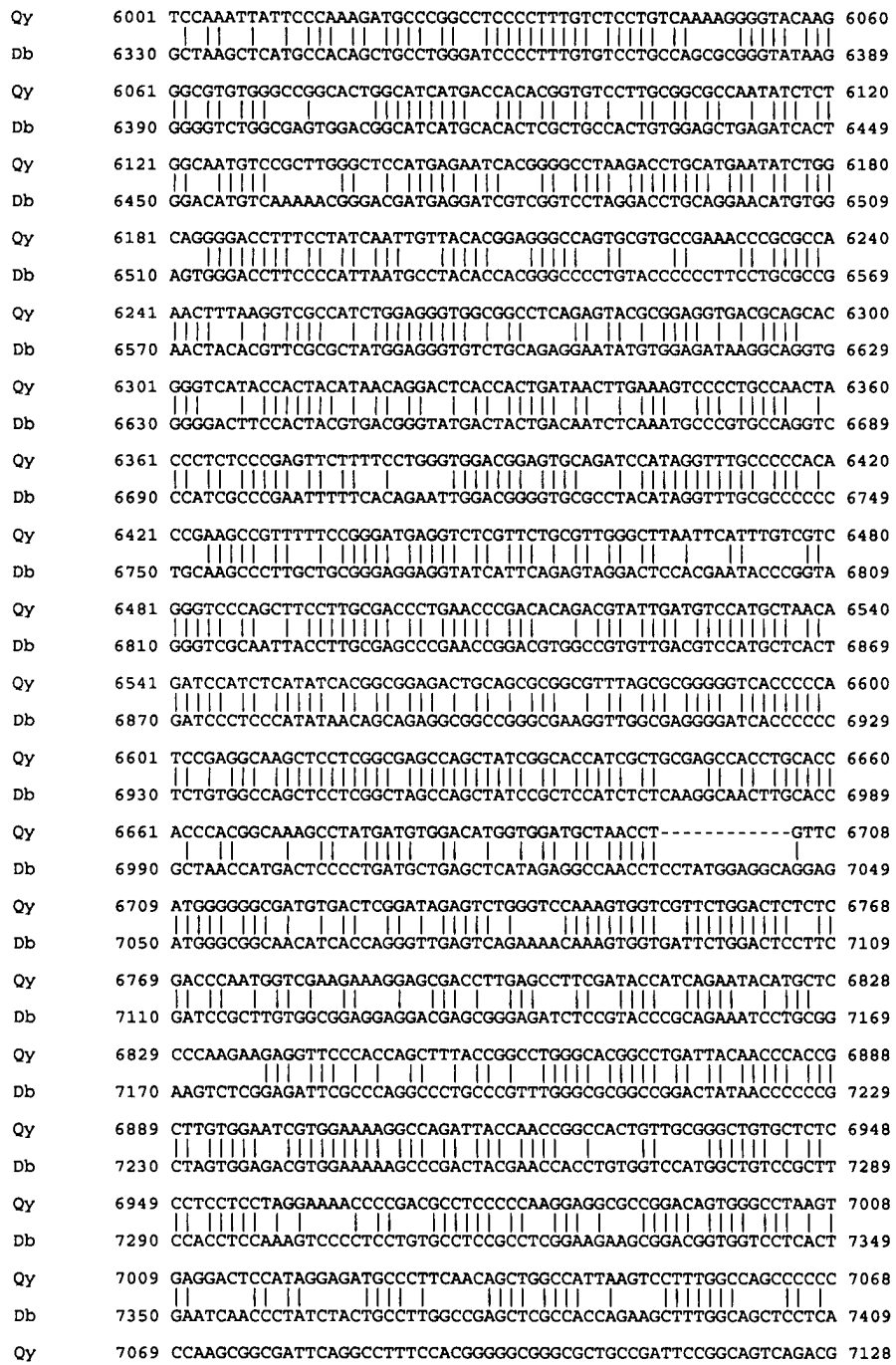
Qy       1441 CCAGAGGATATGAGACCCTATTGCTGGCACTACCCACCAAGGCAGTGTGGCGTGGTCTCC 1500
                ||   | | ||||| |||||||||||||| ||||  |  || ||  | ||  ||
Db       1780 ----CCGACCAGCGCCCCTACTGCTGGCACTACCCCCCAAAACCTTGCGGTATTGTGCCC 1835

Qy       1501 GCGAAGACTGTGTGTGGCCCAGTGTACTGTTTCACCCCCAGCCCAGTGGTAGTGGGCACG 1560
              ||||||| ||||||||| || || || || ||||| |||||||| ||||| ||||| |||
Db       1836 GCGAAGAGTGTGTGTGGTCCGGTATATTGCTTCACTCCCAGCCCCGTGGTGGTGGGAACG 1895

Qy       1561 ACCGACAGGCTTGGAGCGCCCACTTACACGTGGGGGGAGAATGAGACAGATGTCTTCCTA 1620
              |||||||||   || |||||||| ||||  ||||| || ||||| || || |||||| |
Db       1896 ACCGACAGGTCGGGCGCGCCCACCTACAGCTGGGGTGAAAATGATACGGACGTCTTCGTC 1955
```


Qy 1621 TTGAACAGCACTCGACCACCGCTGGGGTCATGGTTCGGCTGCACGTGGATGAACTCTTCT 1680
Db 1956 CTTAACAATACCAGGCCACCGCTGGGCAATTGGTTCGGTTGTACCTGGATGAACTCAACT 2015
Qy 1681 GGCTACACCAAGACTTGCGGCGCACCACCCTGCCGTACTAGAGCTGACTTCAACGCCAGC 1740
Db 2016 GGATTCACCAAAGTGTGCGGAGCGCCTCCTTGTGTCATCGGAGGGGCGGGCAACAACA-- 2073
Qy 1741 ACGGACCTGTTGTGCCCCACGGACTGTTTTAGGAAGCATCCTGATACCACTTACCTCAAA 1800
Db 2074 ----CCCTGCACTGCCCCACTGATTGCTTCCGCAAGCATCCGGACGCCACATACTCTCGG 2129
Qy 1801 TGCGGCTCTGGGCCCTGGCTCACGCCAAGGTGCCTGATCGACTACCCCTACAGGCTCTGG 1860
Db 2130 TGCGGCTCCGGTCCCTGGATCACACCCAGGTGCCTGGTCGACTACCCGTATAGGCTTTGG 2189
Qy 1861 CATTACCCCTGCACAGTTAACTATACCATCTTCAAAATAAGGATGTATGTGGGAGGGGTT 1920
Db 2190 CATTATCCTTGTACCATCAACTACACCATATTTAAAATCAGGATGTACGTGGGAGGGGTC 2249
Qy 1921 GAGCACAGGCTCACGGCTGCATGCAATTTCACTCGTGGGGATCGTTGCAACTTGGAGGAC 1980
Db 2250 GAACACAGGCTGGAAGCTGCCTGCAACTGGACGCGGGGCGAACGTTGCGATCTGGAAGAC 2309
Qy 1981 AGAGACAGAAGTCAACTGTCTCCTTTGTTGCACTCCACCACGGAATGGGCCATTTTACCT 2040
Db 2310 AGGGACAGGTCCGAGCTCAGCCCGTTACTGCTGACCACTACACAGTGGCAGGTCCTCCCG 2369
Qy 2041 TGCTCTTACTCGGACCTGCCCGCCTTGTCGACTGGTCTTCTCCACCTCCACCAAAACATC 2100
Db 2370 TGTTCCTTCACAACCCTACCAGCCTTGTCCACCGGCCTCATCCACCTCCACCAGAACATT 2429
Qy 2101 GTGGACGTACAATTCATGTATGGCCTATCACCTGCCCTCACAAAATACATCGTCCGATGG 2160
Db 2430 GTGGACGTGCAGTACTTGTACGGGGTGGGGTCAAGCATCGCGTCCTGGGCCATTAAGTGG 2489
Qy 2161 GAGTGGGTAATACTCTTATTCCTGCTCTTAGCGGACGCCAGGGTTTGCGCCTGCTTATGG 2220
Db 2490 GAGTACGTCGTTCTCCTGTTCCTTCTGCTTGCAGACGCGCGCGTCTGCTCCTGCTTGTGG 2549
Qy 2221 ATGCTCATCTTGTTGGGCCAGGCCGAAGCAGCACTAGAGAAGCTGGTCATCTTGCACGCT 2280
Db 2550 ATGATGCTACTCATATCCCAAGCGGAGGCGGCTTTGGAGAACCTCGTAATACTTAATGCA 2609
Qy 2281 GCGAGCGCAGCTAGCTGCAATGGCTTCCTATATTTTGTCATCTTTTTCGTGGCTGCTTGG 2340
Db 2610 GCATCCCTGGCCGGGACGCACGGTCTTGTATCCTTCCTCGTGTTCTTCTGCTTTGCATGG 2669
Qy 2341 TACATCAAGGGTCGGGTAGTCCCCTTAGCTACCTATTCCCTCACTGGCCTGTGGTCCTTT 2400
Db 2670 TATTTGAAGGGTAAGTGGGTGCCCGGAGCGGTCTACACCTTCTACGGGATGTGGCCTCTC 2729
Qy 2401 AGCCTACTGCTCCTAGCATTGCCCCAACAGGCTTATGCTTATGACGCATCTGTGCATGGC 2460
Db 2730 CTCCTGCTCCTGTTGGCGTTGCCCCAGCGGGCGTACGCGCTGGACACGGAGGTGGCCGCG 2789
Qy 2461 CAGATAGGAGCGGCTCTGCTGGTAATGATCACTCTCTTTACTCTCACCCCCGGGTATAAG 2520
Db 2790 TCGTGTGGCGGTGTTGTTCTCGTCGGGTTGATGGCGCTGACTCTGTCACCATATTACAAG 2849
Qy 2521 ACCCTTCTCAGCCGGTTTTTGTGGTGGTTGTGCTATCTTCTGACCCTGGGGGAAGCTATG 2580
Db 2850 CGCTATATCAGCTGGTGCTTGTGGTGGCTTCAGTATTTTCTGACCAGAGTGGAAGCGCAA 2909
Qy 2581 GTCCAGGAGTGGGCACCACCTATGCAGGTGCGCGGTGGCCGTGATGGCATCATATGGGCC 2640
Db 2910 CTGCACGTGTGGATTCCCCCCCTCAACGTCCGAGGGGGGCGCGACGCCGTCATCTTACTC 2969
Qy 2641 GTCGCCATATTCTACCCAGGTGTGGTGTTTGACATAACCAAGTGGCTCTTGGCGGTGCTT 2700
Db 2970 ATGTGTGCTGTACACCCGACTCTGGTATTTGACATCACCAAATTGCTGCTGGCCGTCTTC 3029
Qy 2701 GGGCCTGCTTACCTCCTAAAAGGTGCTTTGACGCGCGTGCCGTACTTCGTCAGGGCTCAC 2760
Db 3030 GGACCCCTTTGGATTCTTCAAGCCAGTTTGCTTAAAGTACCCTACTTTGTGCGCGTCCAA 3089
Qy 2761 GCTCTACTGAGGATGTGCACCATGGCAAGGCATCTCGCGGGGGGCAGGTACGTCCAGATG 2820
Db 3090 GGCCTTCTCCGGTTCTGCGCGTTAGCGCGGAAGATGATCGGAGGCCATTACGTGCAAATG 3149
Qy 2821 GCGCTACTAGCCCTTGGCAGGTGGACTGGCACTTACATCTATGACCACCTCACCCCTATG 2880
Db 3150 GTCATCATTAAGTTAGGGGCGCTTACTGGCACCTATGTTTATAACCATCTCACTCCTCTT 3209
Qy 2881 TCGGATTGGGCTGCTAGTGGCCTGCGGGACCTGGCGGTCGCCGTTGAGCCTATCATCTTC 2940
Db 3210 CGGGACTGGGCGCACAACGGCTTGCGAGATCTGGCCGTGGCTGTAGAGCCAGTCGTCTTC 3269
Qy 2941 AGTCCGATGGAGAAGAAAGTCATTGTCTGGGGAGCGGAGACAGCTGCTTGTGGGGACATT 3000
Db 3270 TCCCAAATGGAGACCAAGCTCATCACGTGGGGGGCAGATACCGCCGCGTGCGGTGACATC 3329
Qy 3001 TTACACGGACTTCCCGTGTCCGCCCGACTTGGTCGGGAGGTCCTCCTTGGCCCAGCTGAT 3060
Db 3330 ATCAACGGCTTGCCTGTTTCCGCCCGCAGGGGCCGGGAGATACTGCTCGGGCCAGCCGAT 3389
Qy 3061 GGCTATACCTCCAAGGGGTGGAGTCTTCTCGCCCCCATCACTGCTTACGCCCAGCAGACA 3120
Db 3390 GGAATGGTCTCCAAGGGGTGGAGGTTGCTGGCGCCCATCACGGCGTACGCCCAGCAGACA 3449
Qy 3121 CGTGGCCTTTTGGGCACCATAGTGGTGAGCATGACGGGGCGCGACAAGACAGAACAGGCT 3180
Db 3450 AGGGGCCTCCTAGGGTGCATAATCACCAGCCTAACTGGCCGGGACAAAAACCAAGTGGAG 3509
Qy 3181 GGGGAAATTCAGGTCCTGTCCACAGTCACTCAGTCCTTCCTCGGAACATCCATCTCGGGG 3240
Db 3510 GGTGAGGTCCAGATTGTGTCAACTGCTGCCCAAACCTTCCTGGCAACGTGCATCAATGGG 3569
Qy 3241 GTTTTGTGGACTGTCTACCATGGAGCTGGCAACAAGACTCTGGCCGGCTCACGGGGTCCG 3300
Db 3570 GTGTGCTGGACTGTCTACCACGGGGCCGGAACGAGGACCATCGCGTCACCCAAGGGTCCT 3629
Qy 3301 GTCACGCAGATGTACTCCAGTGCTGAGGGGGACTTAGTAGGGTGGCCCAGCCCCCCTGGG 3360
Db 3630 GTCATCCAGATGTATACCAATGTAGACCAAGACCTTGTGGGCTGGCCCGCTCCGCAAGGT 3689
Qy 3361 ACTAAATCTTTGGAGCCGTGCACGTGTGGAGCGGTCGACCTGTACCTGGTCACGCGGAAC 3420
Db 3690 AGCCGCTCATTGACACCCTGCACTTGCGGCTCCTCGGACCTTTACCTGGTCACGAGGCAC 3749
Qy 3421 GCTGATGTCATCCCGGCTCGAAGACGCGGGGACAAACGGGGAGCGCTACTCTCCCCGAGA 3480
Db 3750 GCCGATGTCATTCCCGTGCGCCGGCGGGGTGATAGCAGGGGCAGCCTGCTGTCGCCCCGG 3809
Qy 3481 CCTCTTTCCACCTTGAAGGGGTCCTCAGGAGGCCCGGTGCTATGCCCCAGGGGCCACGCT 3540
Db 3810 CCCATTTCCTACTTGAAAGGCTCCTCGGGGGGTCCGCTGTTGTGCCCCGCGGGGCACGCC 3869
Qy 3541 GTCGGAGTCTTCCGGGCAGCTGTGTGCTCTCGGGGCGTGGCTAAGTCCATAGATTTCATC 3600
Db 3870 GTGGGCATATTTAGGGCCGCGGTGTGCACCCGTGGAGTGGCTAAGGCGGTGGACTTTATC 3929
Qy 3601 CCCGTTGAGACACTCGACATCGTCACGCGGTCCCCCACCTTTAGTGACAACAGCACACCA 3660
Db 3930 CCTGTGGAGAACCTAGAGACAACCATGAGGTCCCCGGTGTTCACGGATAACTCCTCTCCA 3989
Qy 3661 CCTGCTGTGCCCCAGACCTATCAGGTCGGGTACTTGCATGCCCCGACTGGCAGTGGAAAG 3720
Db 3990 CCAGTAGTGCCCCAGAGCTTCCAGGTGGCTCACCTCCATGCTCCCACAGGCAGCGGCAAA 4049
Qy 3721 AGCACCAAAGTTCCTGTCGCATATGCTGCTCAGGGGTATAAAGTGCTAGTGCTTAATCCC 3780
Db 4050 AGCACCAAGGTCCCGGCTGCATATGCAGCTCAGGGCTATAAGGTGCTAGTACTCAACCCC 4109
Qy 3781 TCAGTGGCTGCCACCCTGGGGTTTGGGGCGTACTTGTCTAAGGCACATGGCATCAATCCC 3840


Db 4110 TCTGTTGCTGCAACACTGGGCTTTGGTGCTTACATGTCCAAGGCTCATGGGATCGATCCT 4169
Qy 3841 AACATTAGGACTGGAGTCAGGACTGTGACGACCGGGGCGCCCATCACGTACTCCACATAT 3900
Db 4170 AACATCAGGACCGGGGTGAGAACAATTACCACTGGCAGCCCCATCACGTACTCCACCTAC 4229
Qy 3901 GGCAAATTCCTCGCCGATGGGGGCTGTGCGGGCGGCGCCTACGACATCATCATATGTGAT 3960
Db 4230 GGCAAGTTCCTTGCCGACGGCGGGTGCTCGGGGGGCGCTTATGACATAATAATTTGTGAC 4289
Qy 3961 GAATGCCATGCCGTGGACTCTACCACCATCCTTGGCATCGGAACAGTCCTTGATCAAGCA 4020
Db 4290 GAGTGCCACTCCACGGATGCCACATCCATCTTGGGCATCGGCACTGTCCTTGACCAAGCA 4349
Qy 4021 GAGACAGCTGGGGTCAGACTAACTGTGCTGGCTACAGCTACGCCCCCTGGGTCAGTGACA 4080
Db 4350 GAGACTGCGGGGGCGAGACTGGTTGTGCTCGCCACCGCCACCCCTCCGGGCTCCGTCACT 4409
Qy 4081 ACCCCCCACCCCAACATAGAGGAGGTGGCCCTTGGGCAGGAGGGCGAGATCCCCTTCTAT 4140
Db 4410 GTGCCCCATCCCAACATCGAGGAGGTTGCTCTGTCCACCACCGGAGAGATCCCTTTTTAC 4469
Qy 4141 GGGAGGGCGATTCCCCTGTCTTACATCAAGGGAGGAAGACATCTGATCTTCTGCCATTCA 4200
Db 4470 GGCAAGGCTATCCCCCTCGAAGTAATCAAGGGGGGGAGACATCTCATCTTCTGTCATTCA 4529
Qy 4201 AAGAAAAAGTGTGACGAGCTCGCGGCGGCCCTTCGGGGTATGGGCTTGAACTCAGTGGCA 4260
Db 4530 AAGAAGAAGTGCGACGAACTCGCCGCAAAGCTGGTCGCATTGGGCATCAATGCCGTGGCC 4589
Qy 4261 TACTACAGAGGGTTGGACGTCTCCGTAATACCAACTCAGGGAGACGTAGTGGTCGTCGCC 4320
Db 4590 TACTACCGCGGTCTTGACGTGTCCGTCATCCCGACCAGCGGCGATGTTGTCGTCGTGGCA 4649
Qy 4321 ACCGACGCCCTCATGACAGGGTATACTGGGGACTTTGACTCCGTGATCGACTGCAACGTA 4380
Db 4650 ACCGATGCCCTCATGACCGGCTATACCGGCGACTTCGACTCGGTGATAGACTGCAATACG 4709
Qy 4381 GCGGTCACTCAAGTTGTAGACTTCAGTTTAGACCCCACATTCACCATAACCACACAGATT 4440
Db 4710 TGTGTCACCCAGACAGTCGATTTCAGCCTTGACCCTACCTTCACCATTGAGACAATCACG 4769
Qy 4441 GTCCCTCAAGACGCTGTCTCACGTAGCCAGCGCCGGGGTCGCACGGGTAGGGGAAGACTG 4500
Db 4770 CTCCCCCAGGATGCTGTCTCCCGCACTCAACGTCGGGGCAGGACTGGCAGGGGGAAGCCA 4829
Qy 4501 GGCATTTATAGGTATGTTTCCACTGGTGAGCGAGCCTCAGGAATGTTTGACAGTGTAGTG 4560
Db 4830 GGCATCTACAGATTTGTGGCACCGGGGGAGCGCCCCTCCGGCATGTTCGACTCGTCCGTC 4889
Qy 4561 CTCTGTGAGTGCTACGACGCAGGGGCCGCATGGTATGAGCTCACACCATCGGAGACCACC 4620
Db 4890 CTCTGTGAGTGCTATGACGCAGGCTGTGCTTGGTATGAGCTCACGCCCGCCGAGACTACA 4949
Qy 4621 GTCAGGCTCAGGGCGTATTTCAACACGCCCGGTTTGCCTGTGTGCCAAGACCATCTTGAG 4680
Db 4950 GTTAGGCTACGAGCGTACATGAACACCCCGGGGCTTCCCGTGTGCCAGGACCATCTTGAA 5009
Qy 4681 TTTTGGGAGGCAGTTTTCACCGGCCTCACACACATAGATGCCCACTTCCTTTCCCAAACA 4740
Db 5010 TTTTGGGAGGGCGTCTTTACAGGCCTCACTCATATAGATGCCCACTTTCTATCCCAGACA 5069
Qy 4741 AAGCAATCGGGGGAAAATTTCGCATACTTAACAGCCTACCAGGCTACAGTGTGCGCTAGG 4800
Db 5070 AAGCAGAGTGGGGAGAACCTTCCTTACCTGGTAGCGTACCAAGCCACCGTGTGCGCTAGG 5129
Qy 4801 GCCAAAGCCCCCCCCCCGTCCTGGGACGTCATGTGGAAGTGTTTGACTCGACTCAAGCCC 4860
Db 5130 GCTCAAGCCCCTCCCCCATCGTGGGACCAGATGTGGAAGTGTTTGATTCGCCTCAAGCCC 5189
Qy 4861 ACACTCGTGGGCCCCACACCTCTCCTGTACCGCTTGGGCTCTGTTACCAACGAGGTCACC 4920
Db 5190 ACCCTCCATGGGCCAACACCCCTGCTATACAGACTGGGCGCTGTTCAGAATGAAATCACC 5249
Qy 4921 CTCACACATCCCGTGACGAAATACATCGCCACCTGCATGCAAGCCGACCTTGAGGTCATG 4980
Db 5250 CTGACGCACCCAGTCACCAAATACATCATGACATGCATGTCGGCCGACCTGGAGGTCGTC 5309
Qy 4981 ACCAGCACATGGGTCTTGGCAGGGGGAGTCTTGGCGGCCGTCGCCGCGTATTGCCTGGCG 5040
Db 5310 ACGAGCACCTGGGTGCTCGTTGGCGGCGTCCTGGCTGCTTTGGCCGCGTATTGCCTGTCA 5369
Qy 5041 ACCGGGTGTGTTTGCATCATCGGCCGCTTGCACATTAACCAGCGAGCCGTCGTTGCGCCG 5100
Db 5370 ACAGGCTGCGTGGTCATAGTGGGCAGGGTCGTCTTGTCCGGGAAGCCGGCAATCATACCT 5429
Qy 5101 GACAAGGAGGTCCTCTATGAGGCTTTTGATGAGATGGAGGAATGTGCCTCTAGGGCGGCT 5160
Db 5430 GACAGGGAAGTCCTCTACCGAGAGTTCGATGAGATGGAAGAGTGCTCTCAGCACTTACCG 5489
Qy 5161 CTCATTGAAGAGGGGCAGCGGATAGCCGAGATGCTGAAGTCCAAGATCCAAGGCTTATTG 5220
Db 5490 TACATCGAGCAAGGGATGATGCTCGCCGAGCAGTTCAAGCAGAAGGCCCTCGGCCTCCTG 5549
Qy 5221 CAGCAAGCTTCCAAACAAGCTCAAGACATACAACCCACTGTGCAGGCTTCATGGCCCAAG 5280
Db 5550 CAGACCGCGTCCCGTCAGGCAGAGGTTATCGCCCCTGCTGTCCAGACCAACTGGCAAAAA 5609
Qy 5281 GTAGAACAATTCTGGGCCAAACACATGTGGAACTTCATTAGCGGCATCCAATACCTCGCA 5340
Db 5610 CTCGAGACCTTCTGGGCGAAGCATATGTGGAACTTCATCAGTGGGATACAATACTTGGCG 5669
Qy 5341 GGACTATCAACACTGCCAGGGAACCCTGCAGTAGCTTCCATGATGGCGTTCAGTGCCGCC 5400
Db 5670 GGCTTGTCAACGCTGCCTGGTAACCCCGCCATTGCTTCATTGATGGCTTTTACAGCTGCT 5729
Qy 5401 CTCACCAGTCCGCTGTCAACAAGCACCACTATCCTTCTCAACATTTTGGGGGGCTGGCTA 5460
Db 5730 GTCACCAGCCCACTAACCACTAGCCAAACCCTCCTCTTCAACATATTGGGGGGGTGGGTG 5789
Qy 5461 GCATCCCAAATTGCACCACCCGCGGGGGCCACTGGCTTCGTTGTCAGTGGCCTAGTGGGA 5520
Db 5790 GCTGCCCAGCTCGCCGCCCCCGGTGCCGCTACTGCCTTTGTGGGCGCTGGCTTAGCTGGC 5849
Qy 5521 GCTGCCGTAGGCAGTATAGGCTTAGGTAAGGTGCTAGTGGACATCCTGGCAGGGTATGGT 5580
Db 5850 GCCGCCATCGGCAGTGTTGGACTGGGGAAGGTCCTCATAGACATCCTTGCAGGGTATGGC 5909
Qy 5581 GCGGGCATTTCGGGGGCTCTCGTCGCATTCAAGATCATGTCTGGCGAGAAGCCCTCCATG 5640
Db 5910 GCGGGCGTGGCGGGAGCTCTTGTGGCATTCAAGATCATGAGCGGTGAGGTCCCCTCCACG 5969
Qy 5641 GAGGATGTCGTCAACTTGCTGCCTGGAATTCTGTCTCCGGGTGCCTTGGTAGTGGGAGTC 5700
Db 5970 GAGGACCTGGTCAATCTACTGCCCGCCATCCTCTCGCCCGGAGCCCTCGTAGTCGGCGTG 6029
Qy 5701 ATCTGCGCGGCCATTCTGCGCCGACACGTGGGACCGGGGGAAGGCGCCGTCCAATGGATG 5760
Db 6030 GTCTGTGCAGCAATACTGCGCCGGCACGTTGGCCCGGGCGAGGGGGCAGTGCAGTGGATG 6089
Qy 5761 AATAGACTCATTGCCTTTGCTTCCAGAGGAAATCACGTCGCCCCCACCCACTACGTGACG 5820
Db 6090 AACCGGCTGATAGCCTTCGCCTCCCGGGGGAACCATGTTTCCCCCACGCACTACGTGCCG 6149
Qy 5821 GAGTCGGATGCGTCGCAGCGTGTGACCCAACTACTTGGCTCCCTTACCATAACCAGCCTG 5880
Db 6150 GAGAGCGATGCAGCTGCCCGCGTCACTGCCATACTCAGCAGCCTCACTGTAACCCAGCTC 6209
Qy 5881 CTCAGAAGACTCCACAACTGGATTACTGAGGACTGCCCCATCCCATGCGGCGGCTCGTGG 5940
Db 6210 CTGAGGCGACTGCACCAGTGGATAAGCTCGGAGTGTACCACTCCATGCTCCGGTTCCTGG 6269
Qy 5941 CTCCGCGATGTGTGGGACTGGGTTTGCACCATCCTAACAGACTTTAAAAATTGGCTGACC 6000
Db 6270 CTAAGGGACATCTGGGACTGGATATGCGAGGTGTTGAGCGACTTTAAGACCTGGCTAAAA 6329


Qy 6001 TCCAAATTATTCCCAAAGATGCCCGGCCTCCCCTTTGTCTCCTGTCAAAAGGGGTACAAG 6060
Db 6330 GCTAAGCTCATGCCACAGCTGCCTGGGATCCCCTTTGTGTCCTGCCAGCGCGGGTATAAG 6389
Qy 6061 GGCGTGTGGGCCGGCACTGGCATCATGACCACACGGTGTCCTTGCGGCGCCAATATCTCT 6120
Db 6390 GGGGTCTGGCGAGTGGACGGCATCATGCACACTCGCTGCCACTGTGGAGCTGAGATCACT 6449
Qy 6121 GGCAATGTCCGCTTGGGCTCCATGAGAATCACGGGGCCTAAGACCTGCATGAATATCTGG 6180
Db 6450 GGACATGTCAAAAACGGGACGATGAGGATCGTCGGTCCTAGGACCTGCAGGAACATGTGG 6509
Qy 6181 CAGGGGACCTTTCCTATCAATTGTTACACGGAGGGCCAGTGCGTGCCGAAACCCGCGCCA 6240
Db 6510 AGTGGGACCTTCCCCATTAATGCCTACACCACGGGCCCCTGTACCCCCCTTCCTGCGCCG 6569
Qy 6241 AACTTTAAGGTCGCCATCTGGAGGGTGGCGGCCTCAGAGTACGCGGAGGTGACGCAGCAC 6300
Db 6570 AACTACACGTTCGCGCTATGGAGGGTGTCTGCAGAGGAATATGTGGAGATAAGGCAGGTG 6629
Qy 6301 GGGTCATACCACTACATAACAGGACTCACCACTGATAACTTGAAAGTCCCCTGCCAACTA 6360
Db 6630 GGGGACTTCCACTACGTGACGGGTATGACTACTGACAATCTCAAATGCCCGTGCCAGGTC 6689
Qy 6361 CCCTCTCCCGAGTTCTTTTCCTGGGTGGACGGAGTGCAGATCCATAGGTTTGCCCCCACA 6420
Db 6690 CCATCGCCCGAATTTTTCACAGAATTGGACGGGGTGCGCCTACATAGGTTTGCGCCCCCC 6749
Qy 6421 CCGAAGCCGTTTTTCCGGGATGAGGTCTCGTTCTGCGTTGGGCTTAATTCATTTGTCGTC 6480
Db 6750 TGCAAGCCCTTGCTGCGGGAGGAGGTATCATTCAGAGTAGGACTCCACGAATACCCGGTA 6809
Qy 6481 GGGTCCCAGCTTCCTTGCGACCCTGAACCCGACACAGACGTATTGATGTCCATGCTAACA 6540
Db 6810 GGGTCGCAATTACCTTGCGAGCCCGAACCGGACGTGGCCGTGTTGACGTCCATGCTCACT 6869
Qy 6541 GATCCATCTCATATCACGGCGGAGACTGCAGCGCGGCGTTTAGCGCGGGGGTCACCCCCA 6600
Db 6870 GATCCCTCCCATATAACAGCAGAGGCGGCCGGGCGAAGGTTGGCGAGGGGATCACCCCCC 6929
Qy 6601 TCCGAGGCAAGCTCCTCGGCGAGCCAGCTATCGGCACCATCGCTGCGAGCCACCTGCACC 6660
Db 6930 TCTGTGGCCAGCTCCTCGGCTAGCCAGCTATCCGCTCCATCTCTCAAGGCAACTTGCACC 6989
Qy 6661 ACCCACGGCAAAGCCTATGATGTGGACATGGTGGATGCTAACCT----------GTTC 6708
Db 6990 GCTAACCATGACTCCCCTGATGCTGAGCTCATAGAGGCCAACCTCCTATGGAGGCAGGAG 7049
Qy 6709 ATGGGGGGCGATGTGACTCGGATAGAGTCTGGGTCCAAAGTGGTCGTTCTGGACTCTCTC 6768
Db 7050 ATGGGCGGCAACATCACCAGGGTTGAGTCAGAAAACAAAGTGGTGATTCTGGACTCCTTC 7109
Qy 6769 GACCCAATGGTCGAAGAAAGGAGCGACCTTGAGCCTTCGATACCATCAGAATACATGCTC 6828
Db 7110 GATCCGCTTGTGGCGGAGGAGGACGAGCGGGAGATCTCCGTACCCGCAGAAATCCTGCGG 7169
Qy 6829 CCCAAGAAGAGGTTCCCACCAGCTTTACCGGCCTGGGCACGGCCTGATTACAACCCACCG 6888
Db 7170 AAGTCTCGGAGATTCGCCCAGGCCCTGCCCGTTTGGGCGCGGCCGGACTATAACCCCCCG 7229
Qy 6889 CTTGTGGAATCGTGGAAAAGGCCAGATTACCAACCGGCCACTGTTGCGGGCTGTGCTCTC 6948
Db 7230 CTAGTGGAGACGTGGAAAAAGCCCGACTACGAACCACCTGTGGTCCATGGCTGTCCGCTT 7289
Qy 6949 CCTCCTCCTAGGAAAACCCCGACGCCTCCCCCAAGGAGGCGCCGGACAGTGGGCCTAAGT 7008
Db 7290 CCACCTCCAAAGTCCCCTCCTGTGCCTCCGCCTCGGAAGAAGCGGACGGTGGTCCTCACT 7349
Qy 7009 GAGGACTCCATAGGAGATGCCCTTCAACAGCTGGCCATTAAGTCCTTTGGCCAGCCCCCC 7068
Db 7350 GAATCAACCCTATCTACTGCCTTGGCCGAGCTCGCCACCAGAAGCTTTGGCAGCTCCTCA 7409
Qy 7069 CCAAGCGGCGATTCAGGCCTTTCCACGGGGGCGGGCGCTGCCGATTCCGGCAGTCAGACG 7128
Db 7410 ACTTCCGGCATTACGGGCGACAATACGACAACATCCTCTGAGCCCGCC---------- 7457
Qy 7129 CCTCCTGATGAGTTGGCCCTTTCGGAGACAGGTTCCATCTCTTCCATGCCCCCCCTCGAG 7188
Db 7458 CCTTCTGGCTGCCCCCCCGACTCCGACGCTGAGTCCTATTCCTCCATGCCCCCCCTGGAG 7517
Qy 7189 GGGGAGCTTGGAGATCCAGACCTGGAGCCTGAGCAGGTAGAGCCCCAACCCCCCCCCCAG 7248
Db 7518 GGGGAGCCTGGGGATCCGGA---------- 7537
Qy 7249 GGGGGGGTGGCAGCTCCCGGCTCGGACTCGGGGTCCTGGTCTACTTGCTCCGAGGAGGAC 7308
Db 7538 ----------TCTTAGCGACGGGTCATGGTCAACGGTCAGTAGTGAGGCCAACGCG 7583
Qy 7309 GACTCCGTCGTGTGCTGCTCCATGTCATACTCCTGGACCGGGGCTCTAATAACTCCTTGT 7368
Db 7584 GAGGATGTCGTGTGCTGCTCAATGTCTTACTCTTGGACAGGCGCACTCGTCACCCCGTGC 7643
Qy 7369 AGTCCCGAAGAGGAGAAGTTACCGATTAACCCCTTGAGCAACTCCCTGTTGCGATATCAC 7428
Db 7644 GCCGCGGAAGAACAGAAACTGCCCATCAATGCACTAAGCAACTCGTTGCTACGTCACCAC 7703
Qy 7429 AACAAGGTGTACTGTACCACAACAAAGAGCGCCTCACTAAGGGCTAAAAAGGTAACTTTT 7488
Db 7704 AATTTGGTGTATTCCACCACCTCACGCAGTGCTTGCCAAAGGCAGAAGAAAGTCACATTT 7763
Qy 7489 GATAGGATGCAAGTGCTCGACTCCTACTACGACTCAGTCTTAAAGGACATTAAGCTAGCG 7548
Db 7764 GACAGACTGCAAGTTCTGGACAGCCATTACCAGGACGTACTCAAGGAGGTTAAAGCAGCG 7823
Qy 7549 GCCTCCAAGGTCACCGCAAGGCTCCTCACCATGGAGGAGGCTTGCCAGTTAACCCCACCC 7608
Db 7824 GCGTCAAAAGTGAAGGCTAACTTGCTATCCGTAGAGGAAGCTTGCAGCCTGACGCCCCCA 7883
Qy 7609 CATTCTGCAAGATCTAAATATGGGTTTGGGGCTAAGGAGGTCCGCAGCTTGTCCGGGAGG 7668
Db 7884 CACTCAGCCAAATCCAAGTTTGGTTATGGGGCAAAAGACGTCCGTTGCCATGCCAGAAAG 7943
Qy 7669 GCCGTTAACCACATCAAGTCCGTGTGGAAGGACCTCCTGGAGGACTCAGAAACACCAATT 7728
Db 7944 GCCGTAACCCACATCAACTCCGTGTGGAAAGACCTTCTGGAAGACAATGTAACACCAATA 8003
Qy 7729 CCCACAACCATTATGGCCAAAAATGAGGTGTTCTGCGTGGACCCCACCAAGGGGGGCAAG 7788
Db 8004 GACACTACCATCATGGCTAAGAACGAGGTTTTCTGCGTTCAGCCTGAGAAGGGGGGTCGT 8063
Qy 7789 AAAGCAGCTCGCCTTATCGTTTACCCTGACCTCGGCGTCAGGGTCTGCGAGAAGATGGCC 7848
Db 8064 AAGCCAGCTCGTCTCATCGTGTTCCCCGATCTGGGCGTGCGCGTGTGCGAAAAGATGGCT 8123
Qy 7849 CTTTATGACATTACACAAAAACTTCCTCAGGCGGTGATGGGGGCTTCTTATGGATTCCAG 7908
Db 8124 TTGTACGACGTGGTTACAAAGCTCCCCTTGGCCGTGATGGGAAGCTCCTACGGATTCCAA 8183
Qy 7909 TATTCCCCCGCTCAGCGGGTAGAGTTTCTCTTGAAAGCATGGGCGGAAAAGAAGGACCCT 7968
Db 8184 TACTCACCAGGACAGCGGGTTGAATTCCTCGTGCAAGCGTGGAAGTCCAAGAAAACCCCA 8243
Qy 7969 ATGGGTTTTTCGTATGATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGG 8028
Db 8244 ATGGGGTTCTCGTATGATACCCGCTGCTTTGACTCCACAGTCACTGAGAGCGACATCCGT 8303
Qy 8029 ACTGAGGAGTCCATATATCGGGCCTGCTCCTTGCCCGAGGAGGCCCACACTGCCATACAC 8088
Db 8304 ACGGAGGAGGCAATCTACCAATGTTGTGACCTCGACCCCCAAGCCCGCGTGGCCATCAAG 8363
Qy 8089 TCGCTAACTGAGAGACTTTACGTGGGAGGGCCTATGTTCAACAGCAAGGGCCAAACCTGC 8148
Db 8364 TCCCTCACCGAGAGGCTTTATGTTGGGGGCCCTCTTACCAATTCAAGGGGGGAGAACTGC 8423
Qy 8149 GGGTACAGGCGTTGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAACACCATCACA 8208

```
Db       8424 GGCTATCGCAGGTGCCGCGCGAGCGGCGTACTGACAACTAGCTGTGGTAACACCCTCACT 8483

Qy       8209 TGCTACGTGAAAGCCTTAGCGGCTTGTAAAGCTGCAGGGATAATCGCGCCCACAATGCTG 8268
             |||||| | || |||   || || |||  ||| |||||| |    |    ||| |||||
Db       8484 TGCTACATCAAGGCCCGGGCAGCCTGTCGAGCCGCAGGGCTCCAGGACTGCACCATGCTC 8543

Qy       8269 GTATGCGGCGATGACTTGGTTGTCATCTCAGAAAGCCAGGGGACCGAGGAGGACGAGCGG 8328
             || || ||||| ||||| || || ||||  ||||||  ||||  | |||||||||| | |
Db       8544 GTGTGTGGCGACGACTTAGTCGTTATCTGTGAAAGCGCGGGGGTCCAGGAGGACGCGGCG 8603

Qy       8329 AACCTGAGAGCCTTCACGGAGGCTATGACCAGGTATTCTGCCCCTCCTGGTGACCCCCCC 8388
             | ||||||||||||||||||||||||||||||| || ||||| ||||| ||||||||
Db       8604 AGCCTGAGAGCCTTCACGGAGGCTATGACCAGGTACTCCGCCCCCCCTGGGGACCCCCCA 8663

Qy       8389 AGACCGGAGTATGATCTGGAGCTGATAACATCTTGCTCCTCAAATGTGTCTGTGGCGCTG 8448
               ||| || || ||  ||||||| |||||||| |||||||| || ||||| || || |
Db       8664 CAACCAGAATACGACTTGGAGCTCATAACATCATGCTCCTCCAACGTGTCAGTCGCCCAC 8723

Qy       8449 GGCCCACAAGGCCGCCGCAGATACTACCTGACCAGAGACCCTACCACTCCAATCGCCCGG 8508
             | |      ||     |    |||||||| ||| | |||||||| || ||  ||||  |
Db       8724 GACGGCGCTGGAAAGAGGGTCTACTACCTCACCCGTGACCCTACAACCCCCCTCGCGAGA 8783

Qy       8509 GCTGCCTGGGAAACAGTTAGACACTCCCCTGTCAATTCATGGCTGGGAAACATCATCCAG 8568
             ||||| ||||| |||| |||||| | || |||||||| ||||| || ||||| |||  |
Db       8784 GCTGCGTGGGAGACAGCAAGACACACTCCAGTCAATTCCTGGCTAGGCAACATAATCATG 8843

Qy       8569 TACGCCCCGACCATATGGGCTCGCATGGTCCTGATGACACACTTCTTCTCCATTCTCATG 8628
             |  ||||| ||  | |||||  | ||| | |||||||| || |||||   | | || ||
Db       8844 TTTGCCCCCACACTGTGGGCGAGGATGATACTGATGACCCATTTCTTTAGCGTCCTTATA 8903

Qy       8629 GCTCAAGACACGCTGGACCAGAACCTCAACTTTGAGATGTACGGAGCGGTGTACTCCGTG 8688
             ||    |||  ||| || |||  |||| | |  ||||| ||||| ||    |||||| |
Db       8904 GCCAGGGACCAGCTTGAACAGGCCCTCGATTGCGAGATCTACGGGGCCTGCTACTCCATA 8963

Qy       8689 AGTCCCTTGGACCTCCCAGCTATAATTGAAAGGTTACATGGGCTTGACGCTTTTTCTCTG 8748
                ||  | || || ||  | || ||| |||| | ||||| ||   ||| ||||| ||
Db       8964 GAACCACTTGATCTACCTCCAATCATTCAAAGACTCCATGGCCTCAGCGCATTTTCACTC 9023

Qy       8749 CACACATACACTCCCCACGAACTGACACGGGTGGCTTCAGCCCTCAGAAAACTTGGGGCG 8808
             ||||  ||| ||||    ||| | |   ||||||| ||  ||||||||||||||||
Db       9024 CACAGTTACTCTCCAGGTGAAATTAATAGGGTGGCCGCATGCCTCAGAAAACTTGGGGTA 9083

Qy       8809 CCACCCCTCAGAGCGTGGAAGAGCCGGGCACGTGCAGTCAGGGCGTCCCTCATCTCCCGT 8868
             || ||| |  |||| ||||    |||||| ||    ||| | ||    ||  |  || |
Db       9084 CCGCCCTTGCGAGCTTGGAGACACCGGGCCCGGAGCGTCCGCGCTAGGCTTCTGGCCAGA 9143

Qy       8869 GGGGGGAGAGCGGCCGTTTGCGGTCGATATCTCTTCAATTGGGCGGTGAAGACCAAGCTC 8928
             || || || || ||| | || ||   || |||||||| ||||| || |  || ||||||
Db       9144 GGAGGCAGGGCTGCCATATGTGGCAAGTACCTCTTCAACTGGGCAGTAAGAACAAAGCTC 9203

Qy       8929 AAACTCACTCCATTGCCGGAAGCGCGCCTCCTGGATTTATCCAGCTGGTTCACCGTCGGC 8988
             |||||||||||| |  |||  ||  |||  ||||| || ||| ||||||||| |  |||
Db       9204 AAACTCACTCCAATAGCGGCCGCTGGCCAGCTGGACTTGTCCGGCTGGTTCACGGCTGGC 9263

Qy       8989 GCCGGCGGGGGCGACATTTATCACAGCGTGTCGCGTGCCCGACCCCGCTTATTGCTCTTT 9048
               | |||||||| |||||||||||||||||||| | |||||| |||||||    |    |||
Db       9264 TACAGCGGGGGAGACATTTATCACAGCGTGTCTCATGCCCGGCCCCGCTGGATCTGGTTT 9323

Qy       9049 GGCCTACTCCTACTTTTTGTAGGGGTAGGCCTTTTCCTACTCCCCGCTCG 9098
              |||||||||| |||  || |||||||||| | | ||| ||||||   ||
Db       9324 TGCCTACTCCTGCTTGCTGCAGGGGTAGGCATCTACCTCCTCCCCAACCG 9373


Search completed: November  1, 2005, 04:24:43
Job time : 1140 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2005, 15:35:06 ; Search time 285 Seconds
(without alignments)
4115.945 Million cell updates/sec

Title: US-09-980-559-2
Perfect score: 16120
Sequence: 1 MSTNPKPQRKTKRNTNRRPQ.....LLFGLLLLFVGVGLFLLPAR 3033

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	16120	100.0	3033	4	AAB59172	Aab59172 Protein e
2	16120	100.0	3033	4	AAB31168	Aab31168 Amino aci
3	16120	100.0	3033	4	AAB30729	Aab30729 Amino aci
4	15824	98.2	3033	2	AAR33538	Aar33538 NANBH vir
5	15645	97.1	2940	7	ADC42917	Adc42917 Hepatitis
6	14943	92.7	3033	5	ABG30688	Abg30688 Human HCV
7	13944	86.5	3033	2	AAR33539	Aar33539 NANBH vir
8	13805	85.6	3033	2	AAR33214	Aar33214 NANBH vir
9	13049	80.9	3015	4	AAB30730	Aab30730 Amino aci
10	13049	80.9	3015	4	AAB30732	Aab30732 Amino aci
11	12925	80.2	3015	4	AAB30733	Aab30733 Amino aci
12	12925	80.2	3015	4	AAB30731	Aab30731 Amino aci
13	12034.5	74.7	3010	5	ABG32459	Abg32459 Hepatitis
14	12033.5	74.6	3010	5	ABG32455	Abg32455 Hepatitis
15	12031.5	74.6	3010	5	ABG32451	Abg32451 Hepatitis
16	12028.5	74.6	3010	5	ABG32453	Abg32453 Hepatitis
17	12027.5	74.6	3010	5	ABG32460	Abg32460 Hepatitis
18	12027.5	74.6	3010	5	ABG32454	Abg32454 Hepatitis
19	12026.5	74.6	3010	5	ABG32461	Abg32461 Hepatitis
20	12025.5	74.6	3010	5	ABG32458	Abg32458 Hepatitis
21	12025.5	74.6	3010	5	ABG32457	Abg32457 Hepatitis
22	12025.5	74.6	3010	5	ABG32452	Abg32452 Hepatitis
23	12022	74.6	3011	5	ABG32456	Abg32456 Hepatitis
24	11998.5	74.4	3010	7	ADF88597	Adf88597 Hepatitis
25	11990.5	74.4	3010	2	AAR68864	Aar68864 Hepatitis
26	11989.5	74.4	3010	2	AAY06423	Aay06423 Non-A, no
27	11986.5	74.4	3010	2	AAR82694	Aar82694 Partial H
28	11984.5	74.3	3010	2	AAR20111	Aar20111 Non-A, no
29	11984.5	74.3	3010	2	AAR20091	Aar20091 Non-A, no
30	11977	74.3	3011	5	AAU84597	Aau84597 HCV polyp
31	11973	74.3	3011	2	AAW77397	Aaw77397 Hepatitis
32	11973	74.3	3011	6	ABP71460	Abp71460 Amino aci
33	11973	74.3	3011	8	ADH79949	Adh79949 E2 HCV en
34	11973	74.3	3011	8	ADL72983	Adl72983 Hepatitis
35	11973	74.3	3012	5	AAU99289	Aau99289 Hepatitis
36	11973	74.3	3012	6	ABU61848	Abu61848 HCV H77 c
37	11966	74.2	3011	4	AAB31169	Aab31169 Amino aci
38	11966	74.2	3011	5	AAU79221	Aau79221 Hepatitis
39	11966	74.2	3011	8	ADJ56744	Adj56744 Hepatitis
40	11966	74.2	3011	8	ADJ64256	Adj64256 Hepatitis
41	11966	74.2	3180	6	ABG73195	Abg73195 MKO-Z vir
42	11964	74.2	3011	2	AAW98020	Aaw98020 Infectiou
43	11963.5	74.2	3010	2	AAR68622	Aar68622 HCV prote
44	11958.5	74.2	3010	2	AAW98022	Aaw98022 Infectiou
45	11958.5	74.2	3010	4	AAB31170	Aab31170 Amino aci

ALIGNMENTS

RESULT 1
AAB59172
ID AAB59172 standard; protein; 3033 AA.
XX
AC AAB59172;
XX
DT 21-MAR-2001 (first entry)
XX
DE Protein encoded by infectious Hepatitis C virus 2a genotype.
XX
KW GBV-B; hepatitis C virus; HCV; vaccine.
XX
OS Hepatitis C virus.
XX
PN WO200075337-A1.
XX
PD 14-DEC-2000.
XX
PF 02-JUN-2000; 2000WO-US015293.
XX
PR 04-JUN-1999; 99US-0137694P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Bukh J, Yanagi M, Emerson SU, Purcell RH;
XX
DR WPI; 2001-091214/10.
XX
PT New infectious nucleic acids of the GB virus-B clone, useful for
PT indirectly studying the molecular properties of hepatitis C virus (HCV)
PT and in developing vaccines and therapeutics for HCV.
XX
PS Disclosure; Page 82-94; 96pp; English.
XX
CC The present invention relates to GB virus-B. The nucleic acid molecules
CC of the invention are useful for indirectly studying the molecular
CC properties of hepatitis C virus (HCV). The infectious nucleic acid
CC sequence of the GB virus-B clone and the HCV/GBV-B chimeras may be used
CC in the development of vaccines and therapeutics for HCV
XX
SQ Sequence 3033 AA;

Query Match 100.0%; Score 16120; DB 4; Length 3033;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3033; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780

Qy        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840

Qy        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900

Qy        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960

Qy        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020

Qy       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080

Qy       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140

Qy       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200

Qy       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260

Qy       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320

Qy       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380

Qy       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440

Qy       1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500

Qy       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560

Qy       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620

Qy       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680

Qy       1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740

Qy       1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800

Qy       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860

Qy       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920

Qy       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980

Qy       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040

Qy       2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100

Qy       2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160

Qy       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220
```

```
Qy       2221 THGKAYDVDMVDANLFMGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYMLPKKR 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 THGKAYDVDMVDANLFMGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYMLPKKR 2280

Qy       2281 FPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLSEDSI 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 FPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLSEDSI 2340

Qy       2341 GDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTPPDELALSETGSISSMPPLEGELG 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTPPDELALSETGSISSMPPLEGELG 2400

Qy       2401 DPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 DPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460

Qy       2461 EKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIKLAASKV 2520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 EKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIKLAASKV 2520

Qy       2521 TARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSETPIPTTI 2580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2521 TARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSETPIPTTI 2580

Qy       2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPA 2640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPA 2640

Qy       2641 QRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTE 2700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2641 QRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTE 2700

Qy       2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPTMLVCGD 2760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPTMLVCGD 2760

Qy       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820

Qy       2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDT 2880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDT 2880

Qy       2881 LDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2881 LDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2940

Qy       2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000

Qy       3001 DIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
             |||||||||||||||||||||||||||||||||
Db       3001 DIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
```

RESULT 2
AAB31168
ID AAB31168 standard; protein; 3033 AA.
XX
AC AAB31168;
XX
DT 02-APR-2001 (first entry)
XX
DE Amino acid sequence of a hepatitis C virus (HCV) clone genotype 2a.
XX
KW Chimeric virus; bovine viral diarrhoea virus; BVDV; hepatitis C virus;
KW HCV; vaccine; viral inhibitor; antiviral.
XX
OS Hepatitis C virus.
XX
PN WO200075352-A2.
XX
PD 14-DEC-2000.
XX
PF 02-JUN-2000; 2000WO-US015527.
XX
PR 04-JUN-1999; 99US-0137817P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Nam J, Bukh J, Emerson SU, Purcell RH;
XX
DR WPI; 2001-071081/08.
DR N-PSDB; AAC86937.
XX
PT New nucleic acid comprising a chimeric bovine viral diarrhea virus genome
PT in which the (non-)structural region has been replaced by hepatitis C
PT virus (HCV) genome useful for treating or preventing HCV signs and
PT symptoms.
XX
PS Disclosure; Page 85-97; 97pp; English.
XX
CC The specification describes a nucleic acid comprising a chimeric virus
CC genome, specifically bovine viral diarrhoea virus (BVDV) genome in which
CC the (non-)structural region has been replaced by the (non-)structural
CC region of a hepatitis C virus (HCV) genome. The nucleic acids comprising
CC the chimeric virus and the chimeric virus are useful for identifying cell
CC lines capable of supporting the replication of these chimeric viruses, in
CC screening for neutralizing antibodies to HCV of different genotypes, in
CC the production of HCV-BVDV virions, for the development of inactivated or
CC attenuated vaccines to prevent HCV-BVDV in a mammal, in studying the
CC molecular properties of HCV indirectly in vitro, and in identifying
CC inhibitors of viral enzyme activity which would be useful as antiviral
CC agents. Formulations or compositions comprising the chimeric virions may
CC be used to treat or prevent the signs and symptoms of HCV. The present
CC sequence is encoded by a HCV clone, which is used to construct chimeric
CC nucleic acids of the invention
XX
SQ Sequence 3033 AA;

Query Match 100.0%; Score 16120; DB 4; Length 3033;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3033; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780

Qy        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840

Qy        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900

Qy        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960

Qy        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020

Qy       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080

Qy       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140

Qy       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200

Qy       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260

Qy       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320

Qy       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380

Qy       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440

Qy       1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTIITTQIVPQDAVSRSQRRGRTGRGRL 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTIITTQIVPQDAVSRSQRRGRTGRGRL 1500

Qy       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560

Qy       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620

Qy       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680

Qy       1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740

Qy       1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800

Qy       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860

Qy       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920

Qy       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980

Qy       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040

Qy       2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100

Qy       2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160

Qy       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220

Qy       2221 THGKAYDVDMVDANLFMGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYMLPKKR 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 THGKAYDVDMVDANLFMGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYMLPKKR 2280

Qy       2281 FPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLSEDSI 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 FPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLSEDSI 2340

Qy       2341 GDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTPPDELALSETGSISSMPPLEGELG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTPPDELALSETGSISSMPPLEGELG 2400

Qy       2401 DPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 DPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460

Qy       2461 EKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIKLAASKV 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 EKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIKLAASKV 2520

Qy       2521 TARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSETPIPTTI 2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2521 TARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSETPIPTTI 2580

Qy       2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPA 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPA 2640
```

```
Qy       2641 QRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTE 2700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2641 QRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTE 2700

Qy       2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPTMLVCGD 2760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPTMLVCGD 2760

Qy       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820

Qy       2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDT 2880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDT 2880

Qy       2881 LDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2881 LDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2940

Qy       2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000

Qy       3001 DIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
             |||||||||||||||||||||||||||||||||
Db       3001 DIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
```

RESULT 3
AAB30729
ID AAB30729 standard; protein; 3033 AA.
XX
AC AAB30729;
XX
DT 02-APR-2001 (first entry)
XX
DE Amino acid sequence of infectious Hepatitis C virus strain HC-J6CH.
XX
KW HCV; HCV strain HC-J6CH; HCV genotype 2a; antiviral; vaccine.
XX
OS Hepatitis C virus.
XX
PN WO200075338-A2.
XX
PD 14-DEC-2000.
XX
PF 02-JUN-2000; 2000WO-US015446.
XX
PR 04-JUN-1999; 99US-0137693P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Yanagi M, Bukh J, Emerson SU, Purcell RH;
XX
DR WPI; 2001-061728/07.
DR N-PSDB; AAC86644.
XX
PT Nucleic acid molecule encoding human hepatitis C virus of genotype 2a for
PT developing vaccines, for diagnosis of hepatitis C virus and in screening
PT assays for identification of antiviral agents.
XX
PS Disclosure; Page 88-99; 167pp; English.
XX
CC The present sequence represents an amino acid sequence of infectious
CC Hepatitis C virus (HCV) strain HC-J6CH genotype 2a. The HCV
CC polynucleotide sequence is capable of expressing the virus when
CC transfected into cells. The HCV protein is useful for assaying candidate
CC antiviral agents for activity against HCV. Antibodies specific for HCV
CC polypeptide are useful in prevention and treatment of diseases caused by
CC HCV in animals, in particular humans. The HCV polypeptides serve as
CC immunogens in the development of vaccines for preventing HCV in mammals
CC or as antigens in diagnostic assays for detecting the presence of HCV in
CC biological samples. The HCV polynucleotide is also useful for identifying
CC cell lines capable of supporting the replication of HCV in vitro and to
CC produce attenuated viral strains via passage in vitro or in vivo
XX
SQ Sequence 3033 AA;

Query Match 100.0%; Score 16120; DB 4; Length 3033;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3033; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780

Qy        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840

Qy        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900

Qy        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
```

```
Qy       961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020

Qy      1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080

Qy      1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140

Qy      1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200

Qy      1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260

Qy      1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320

Qy      1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380

Qy      1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440

Qy      1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500

Qy      1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560

Qy      1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620

Qy      1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680

Qy      1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740

Qy      1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800

Qy      1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860

Qy      1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920

Qy      1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980

Qy      1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040

Qy      2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100

Qy      2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160

Qy      2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220

Qy      2221 THGKAYDVDMVDANLFMGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYMLPKKR 2280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2221 THGKAYDVDMVDANLFMGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYMLPKKR 2280

Qy      2281 FPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLSEDSI 2340
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2281 FPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLSEDSI 2340

Qy      2341 GDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTPPDELALSETGSISSMPPLEGELG 2400
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2341 GDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTPPDELALSETGSISSMPPLEGELG 2400

Qy      2401 DPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2401 DPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460

Qy      2461 EKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIKLAASKV 2520
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2461 EKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIKLAASKV 2520

Qy      2521 TARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSETPIPTTI 2580
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2521 TARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSETPIPTTI 2580

Qy      2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPA 2640
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPA 2640

Qy      2641 QRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTE 2700
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2641 QRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTE 2700

Qy      2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPTMLVCGD 2760
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPTMLVCGD 2760

Qy      2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820

Qy      2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDT 2880
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDT 2880

Qy      2881 LDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2940
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2881 LDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2940

Qy      2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000

Qy      3001 DIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
            |||||||||||||||||||||||||||||||||
Db      3001 DIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033


RESULT 4
AAR33538
ID   AAR33538 standard; protein; 3033 AA.
XX
```

```
AC   AAR33538;
XX
DT   27-AUG-2003  (revised)
DT   25-MAR-2003  (revised)
DT   01-JUL-1993  (first entry)
XX
DE   NANBH virus strain HC-J6 protein.
XX
KW   Non A non B hepatitis virus; plasma.
XX
OS   Non-A.
OS   non-B hepatitis virus.
XX
PN   EP532167-A2.
XX
PD   17-MAR-1993.
XX
PF   30-JUL-1992;   92EP-00306952.
XX
PR   09-AUG-1991;   91JP-00287402.
PR   05-DEC-1991;   91JP-00360441.
XX
PA   (IMMO ) IMMUNO JAPAN INC.
XX
PI   Okamoto H,  Nakamura T;
XX
DR   WPI; 1993-087166/11.
DR   P-PSDB; AAQ38218.
XX
PT   Polynucleotide(s), polypeptide(s) and antibodies of NANBH virus - useful
PT   for detecting NANBH, as a vaccine and for screening blood samples.
XX
PS   Claim 5; Page 38-52; 93pp; English.
XX
CC   RNA was isolated from the plasma of human patients positive for NANBH
CC   virus (strain HC-J6) and was subjected to reverse transcription to
CC   produce cDNA. The resulting cDNA was amplified by PCR, and nucleic acid
CC   sequences determined by analysis of both clones from the cDNA library and
CC   clones obtd. by PCR amplification (36 clones in total). The NANBH HC-J6
CC   genome was found to contain an open reading frame encoding a polypeptide
CC   precursor of 3033 amino acid residues. See also AAR33539 and AAR33214.
CC   (Updated on 25-MAR-2003 to correct PN field.) (Updated on 27-AUG-2003 to
CC   correct OS field.)
XX
SQ   Sequence 3033 AA;

  Query Match             98.2%;  Score 15824;  DB 2;  Length 3033;
  Best Local Similarity   97.8%;  Pred. No. 0;
  Matches 2966;  Conservative   30;  Mismatches   37;  Indels    0;  Gaps    0;

Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| :|||
Db        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNTSRCWI 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
              |||||||||:||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PVSPNVAVQQPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIIGGAHWGVMFGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
              |||||||||||||||||||||||:||||||| |      || :| :| |||||:||||||
Db        361 YFSMQGAWAKVVVILLLAAGVDAQTHTVGGSTAHNARTLTGMFSLGARQKIQLINTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
              ||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
Db        421 HINRTALNCNDSLHTGFLASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
              |||||||||||||||||||| :|||||||||||||||||||||||||||||||||||||
Db        481 PEDMRPYCWHYPPRQCGVVSASSVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
              ||||||| ||||||||||:||||||||||||| ||||||| ||||||||||||||||:|
Db        541 LNSTRPPQGSWFGCTWMNSTGYTKTCGAPPCRIRADFNASMDLLCPTDCFRKHPDTTYIK 600

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
              |||||||||||||||||||||:||||||||||||||||||||||||||||||||||||||
Db        661 RDRSQLSPLLHSTTEWAILPCTYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
              |||:||||||||||||||||||||||||||||:||||||||||||||||||||||||||
Db        721 EWVVLLFLLLADARVCACLWMLILLGQAEAALEKLVVLHAASAASCNGFLYFVIFFVAAW 780

Qy        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
              |||||||||||||||||| ||||||||||||||||||||||||:||||||||||||||||
Db        781 YIKGRVVPLATYSLTGLWSFGLLLLALPQQAYAYDASVHGQIGAALLVLITLFTLTPGYK 840

Qy        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
              |||||||||||||||||| ||||||||||||||||||||||||||| ||||||||||||||
Db        841 TLLSRFLWWLCYLLTLAEAMVQEWAPPMQVRGGRDGIIWAVAIFCPGVVFDITKWLLAVL 900

Qy        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
              |||||||||||||||||||||||||||||| |||||||||| ||||||||||||||||||
Db        901 GPAYLLKGALTRVPYFVRAHALLRMCTMVRHLAGGRYVQMVLLALGRWTGTYIYDHLTPM 960

Qy        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
              |||||:||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SDWAANGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020

Qy       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Db       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTTISG 1080

Qy       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140

Qy       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200

Qy       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260

Qy       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320

Qy       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380

Qy       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440
              |||||||||||||||||||||||||||||||||||||:|||||||||||||||||||||
Db       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNAVAYYRGLDVSVIPTQGDVVVVA 1440
```

```
Qy      1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500
             |||||||:|||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db      1441 TDALMTGFTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQTVPQDAVSRSQRRGRTGRGRL 1500

Qy      1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560
             |||||||||||||||||||||||||||||||||||:||||||||||||||||||||||||
Db      1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPAETTVRLRAYFNTPGLPVCQDHLE 1560

Qy      1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620

Qy      1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680

Qy      1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740
             ||||||||||:||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1681 TGCVCIIGRLHVNQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740

Qy      1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
             |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db      1741 QQASKQAQDIQPAVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800

Qy      1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860

Qy      1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920

Qy      1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db      1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCSGSW 1980

Qy      1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040
             ||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||||
Db      1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFISCQKGYKGVWAGTGIMTTRCPCGANIS 2040

Qy      2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100
             ||||||||||||||||||||||||||||||||||||||||||||:||||||||||||||||
Db      2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKIAIWRVAASEYAEVTQH 2100

Qy      2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160
             |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db      2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPIPKPFFRDEVSFCVGLNSFVV 2160

Qy      2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220
             |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db      2161 GSQLPCDPEPDTDVLTSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220

Qy      2221 THGKAYDVDMVDANLFMGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYMLPKKR 2280
             |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db      2221 THGKAYDVDMVDANLFMGGDVTRIESESKVVVLDSLDPMVEERSDLEPSIPSEYMLPKKR 2280

Qy      2281 FPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLSEDSI 2340
             |||||||||||||||||||||||||||||||||||:|||||||||||||||| ||||||||
Db      2281 FPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPKKTPTPPPRRRRTVGLSESSI 2340

Qy      2341 GDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTPPDELALSETGSISSMPPLEGELG 2400
              |||||||||||||||||||||||| |||||:|||||||||||||||||||||||||| |
Db      2341 ADALQQLAIKSFGQPPPSGDSGLSTGADAADSGSRTPPDELALSETGSISSMPPLEGEPG 2400

Qy      2401 DPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460
             |||||||||| ||||||||  ||| ||||||||||||||||||||||||||||||||||||
Db      2401 DPDLEPEQVELQPPPQGGVVTPGSGSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460

Qy      2461 EKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIKLAASKV 2520
             ||||| |||||||||||||||:|||||||||||||| ||::|||||||||||||||||||
Db      2461 EKLPIGPLSNSLLRYHNKVYCTTSKSASLRAKKVTFDRMQALDAHYDSVLKDIKLAASKV 2520

Qy      2521 TARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSETPIPTTI 2580
             ||||||:|||||||||||||||||||||||||||||||||||||||||||::|||||||
Db      2521 TARLLTLEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDTQTPIPTTI 2580

Qy      2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPA 2640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPA 2640

Qy      2641 QRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTE 2700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2641 QRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTE 2700

Qy      2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPTMLVCGD 2760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPTMLVCGD 2760

Qy      2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820

Qy      2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDT 2880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDT 2880

Qy      2881 LDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2881 LDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2940

Qy      2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000

Qy      3001 DIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
             ||||||||||||||| ||||||||||||||||||
Db      3001 DIYHSVSRARPRLLLLGLLLLFVGVGLFLLPAR 3033
```

RESULT 5
ADC42917
ID ADC42917 standard; protein; 2940 AA.
XX
AC ADC42917;
XX
DT 18-DEC-2003 (first entry)
XX
DE Hepatitis C virus protein.
XX
KW viral; YPXL; cellular protein AP-50; hepatitis C virus; HCV;
KW herpes simplex virus; HSV; UL42 protein; variola virus A10L;
KW vaccinia virus virion core protein P4a;
KW human parainfluenza virus haemagglutinin-neuraminidase; HSP-1;
KW viral budding; viral infectivity.
XX
OS Hepatitis C virus.
XX
PN WO2003017943-A2.
XX
PD 06-MAR-2003.
XX
PF 22-AUG-2002; 2002WO-US027066.
XX
PR 22-AUG-2001; 2001US-0314182P.
XX
PA (MYRI-) MYRIAD GENETICS INC.
XX
PI Morham S, Zavitz K, Hobden A;
XX
DR WPI; 2003-371696/35.
XX
PT Novel peptide, comprising a contiguous amino acid sequence of a viral
PT protein, capable of binding to a region of cellular protein AP-50 is
PT useful for treating viral infections.
XX
PS Example 3; SEQ ID NO 13; 60pp; English.

XX
CC The invention relates to a novel isolated peptide comprising a contiguous
CC amino acid sequence of 7-30 amino acid residues of a viral protein, where
CC the contiguous amino acid sequence encompasses the YPXL motif of the
CC viral protein, and the 7-30 amino acid viral protein is capable of
CC binding a region including the amino acid residues 121-435 of cellular
CC protein AP-50. The isolated 7-30 amino acid viral protein is selected
CC from hepatitis C virus (HCV) polyprotein, herpes simplex virus (HSV) UL42
CC protein, variola virus A10L protein, vaccinia virus virion core protein
CC P4a, and human parainfluenza virus haemagglutinin-neuraminidase, where
CC the contiguous amino acid sequence encompass the YPXL motif of the viral
CC protein. The peptides of the invention are useful in the manufacture of a
CC medicament for treating viral infections caused by a virus selected from
CC HCV, HSP-1 and variola virus. The peptides are also useful for inhibiting
CC viral budding from virus-infected cells and/or inhibiting viral
CC infectivity, thus inhibiting viral propagation in the cells. This
CC sequence represents the protein of a Hepatitis C virus used in the
CC treatment of viral infections of the invention.
XX
SQ Sequence 2940 AA;

Query Match 97.1%; Score 15645; DB 7; Length 2940;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2940; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780

Qy        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840

Qy        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900

Qy        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960

Qy        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020

Qy       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080

Qy       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140

Qy       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200

Qy       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260

Qy       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320

Qy       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380

Qy       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440

Qy       1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500

Qy       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560

Qy       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620

Qy       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680

Qy       1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740

Qy       1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800

Qy       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860
```

```
Db       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860

Qy       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920

Qy       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980

Qy       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040

Qy       2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100

Qy       2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160

Qy       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220

Qy       2221 THGKAYDVDMVDANLFMGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYMLPKKR 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 THGKAYDVDMVDANLFMGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYMLPKKR 2280

Qy       2281 FPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLSEDSI 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 FPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLSEDSI 2340

Qy       2341 GDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTPPDELALSETGSISSMPPLEGELG 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTPPDELALSETGSISSMPPLEGELG 2400

Qy       2401 DPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 DPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460

Qy       2461 EKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIKLAASKV 2520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 EKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIKLAASKV 2520

Qy       2521 TARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSETPIPTTI 2580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2521 TARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSETPIPTTI 2580

Qy       2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPA 2640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPA 2640

Qy       2641 QRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTE 2700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2641 QRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTE 2700

Qy       2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPTMLVCGD 2760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPTMLVCGD 2760

Qy       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820

Qy       2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDT 2880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDT 2880

Qy       2881 LDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2881 LDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2940
```

RESULT 6
ABG30688
ID ABG30688 standard; protein; 3033 AA.
XX
AC ABG30688;
XX
DT 07-OCT-2002 (first entry)
XX
DE Human HCV-related polypeptide.
XX
KW Hepatitis C virus; human; virucide; gene therapy; HCV;
KW fulminant hepatitis C.
XX
OS Homo sapiens.
XX
PN JP2002171978-A.
XX
PD 18-JUN-2002.
XX
PF 01-DEC-2000; 2000JP-00367365.
XX
PR 01-DEC-2000; 2000JP-00367365.
XX
PA (TOKR-) ZH TOKYOTO RINSHO IGAKU SOGO KENKYUSHO.
PA (TORA) TORAY IND INC.
XX
DR WPI; 2002-569884/61.
DR N-PSDB; ABK88904.
XX
PT A gene of a fulminant hepatitis C virus strain and the encoded
PT polypeptide useful in gene therapy to treat hepatitis C.
XX
PS Claim 1; Page 25-33; 36pp; Japanese.
XX
CC The invention relates to a human polypeptide related to hepatitis C virus
CC (HCV), and the polynucleotide encoding it. The polypeptide can be used
CC for the development of gene therapy on fulminant hepatitis C. This
CC sequence represents a human HCV-related polypeptide
XX
SQ Sequence 3033 AA;

Query Match 92.7%; Score 14943; DB 5; Length 3033;
Best Local Similarity 91.2%; Pred. No. 0;
Matches 2765; Conservative 129; Mismatches 139; Indels 0; Gaps 0;

```
Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              |||||||||||||||||:|||||||||||||||||||||||||| ||||||||||||
Db          1 MSTNPKPQRKTKRNTNRRPEDVKFPGGGQIVGGVYLLPRRGPRLGVRTTRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              ||||||||||||:|||| ||||||||||||||||||||||||| ||||||||||
Db         61 RRQPIPKDRRSTGKAWGKPGRPWPLYGNEGLGWAGWLLSPRGSRPSWGPTDPRHRSRNVG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
              |||||||||||||||||||| | |||:||||||||||||:|||||||  |||||||
Db        121 KVIDTLTCGFADLMGYIPVVGAPLSGAARAVAHGVRVLEDGVNYATGNLPGFPFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
              |||||| |||||:||| |: |||||||:||||||||:||||||||||||:||| |:||:
Db        181 LLSCITVPVSAAQVKNTSSSYMVTNDCSNDSITWQLEAAVLHVPGCVPCERVGNTSRCWV 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
              |||||:||::|||||||||||||||||| ||||||||||||||||:||||||:|||
Db        241 PVSPNMAVRQPGALTQGLRTHIDMVVMSATFCSALYVGDLCGGVMLAAQVFIVSPQYHWF 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
              ||:|||||||||||||||||||||||||||||| ||||||||:||||||||||
Db        301 VQECNCSIYPGTITGHRMAWDMMMNWSPTATMILAYVMRVPEVIIDIVSGAHWGVMFGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
```

```
Db        361 YFSMQGAWAKVIVILLLAAGVDAGTTTVGGAVARSTNVIAGVFSHGPQQNIQLINTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480

Db        421 HINRTALNCNDSLNTGFLAALFYTNRFNSSGCPGRLSACRNIEAFRIGWGTLQYEDNVTN 480

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540

Db        481 PEDMRPYCWHYPPKPCGVVPARSVCGPVYCFTPSPVVVGTTDRRGVPTYTWGENETDVFL 540

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600

Db        541 LNSTRPPQGSWFGCTWMNSTGFTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDATYIK 600

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660

Db        601 CGSGPWLTPKCLVHYPYRLWHYPCTVNFTIFKIRMYVGGVEHRLTAACNFTRGDRCDLED 660

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720

Db        661 RDRSQLSPLLHSTTEWAILPCTYSDLPALSTGLLHLHQNIVDVQYMYGLSPAITKYVVRW 720

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780

Db        721 EWVVLLFLLLADARVCACLWMLILLGQAEAALEKLVVLHAASAANCHGLLYFAIFFVAAW 780

Qy        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840

Db        781 HIRGRVVPLTTYCLTGLWPFCLLLMALPRQAYAYDAPVHGQIGVGLLILITLFTLTPGYK 840

Qy        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900

Db        841 TLLGQCLWWLCYLLTLGEAMIQEWVPPMQVRGGRDGIAWAVTIFCPGVVFDITKWLLALL 900

Qy        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960

Db        901 GPAYLLRAALTHVPYFVRAHALIRVCALVKQLAGGRYVQVALLALGRWTGTYIYDHLTPM 960

Qy        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020

Db        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGQEILLGPAD 1020

Qy       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080

Db       1021 GYTSKGWKLLAPITAYAQQTRGLLGAIVVSMTGRDRTEQAGEVQILSTVSQSFLGTTISG 1080

Qy       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140

Db       1081 VLWTVYHGAGNKTLAGLRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCKCGAVDLYLVTRN 1140

Qy       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200

Db       1141 ADVIPARRRGDKRGALLSPRPISTLKGSSGGPVLCPRGHVVGLFRAAVCSRGVAKSIDFI 1200

Qy       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260

Db       1201 PVETLDVVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260

Qy       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320

Db       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVMTGEAITYSTYGKFLADGGCASGAYDIIICD 1320

Qy       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380

Db       1321 ECHAVDATSILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPDIEEVGLGREGEIPFY 1380

Qy       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440

Db       1381 GRAIPLSCIKGGRHLIFCHSKKKCDELAAALRGMGLNAVAYYRGLDVSIIPAQGDVVVVA 1440

Qy       1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500

Db       1441 TDALMTGYTGDFDSVIDCNVAVTQAVDFSLDPTFTITTQTVPQDAVSRSQRRGRTGRGRQ 1500

Qy       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560

Db       1501 GTYRYVSTGERASGMFDSVVLCECYDAGAAWYDLTPAETTVRLRAYFNTPGLPVCQDHLE 1560

Qy       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620

Db       1561 FWEAVFTGLTHIDAHFLSQTKQAGENFAYLVAYQATVCARAKAPPPSWDAMWKCLARLKP 1620

Qy       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680

Db       1621 TLAGPTPLLYRLGPITNEVTLTHPGTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680

Qy       1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740

Db       1681 TGCVSIIGRLHVNQRVVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740

Qy       1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800

Db       1741 QQASKQAQDIQPAMQASWPKVEQFWARHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800

Qy       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860

Db       1801 LTSPLSTSTTILLNIMGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860

Qy       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920

Db       1861 AGISGALVAFKIMSGEKPSMEDVINLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920

Qy       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980

Db       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCSGSW 1980

Qy       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040

Db       1981 LRDVWDWVCTILTDFKNWLTSKLFPKLPGLPFISCQKGYKGVWAGTGIMTTRCPCGANIS 2040

Qy       2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100

Db       2041 GNVRLGSMRITGPKTCMNTWQGTFPINCYTEGQCAPKPPTNYKTAIWRVAASEYAEVTQH 2100

Qy       2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160

Db       2101 GSYSYVTGLTTDNLKIPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSYAV 2160

Qy       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220

Db       2161 GSQLPCEPEPDADVLRSMLTDPPHITAETAARRLARGSPPSEASSSVSQLSAPSLRATCT 2220

Qy       2221 THGKAYDVDMVDANLFMGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYMLPKKR 2280

Db       2221 THSNTYDVDMVDANLLMEGGVAQTEPESRVPVLDFLEPMAEEESDLEPSIPSECMLPRSG 2280

Qy       2281 FPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLSEDSI 2340

Db       2281 FPRALPAWARPDYNPPLVESWRRPDYQPPTVAGCALPPPKKAPTPPPRRRRTVGLSESTI 2340

Qy       2341 GDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTPPDELALSETGSISSMPPLEGELG 2400

Db       2341 SEALQQLAIKTFGQPPSSGDAGSSTGAGAAESGGPTSPGEPAPSETGSASSMPPLEGEPG 2400

Qy       2401 DPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460

Db       2401 DPDLESDQVELQPPPQGGGVAPGSGSGSWSTCSEEDDTTVCCSMSYSWTGALITPCSPEE 2460

Qy       2461 EKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIKLAASKV 2520

Db       2461 EKLPINPLSNSLLRYHNKVYCTTSKSASQRAKKVTFDRTQVLDAHYDSVLKDIKLAASKV 2520

Qy       2521 TARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSETPIPTTI 2580

Db       2521 SARLLTLEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDPQTPIPTTI 2580
```

```
Qy       2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPA 2640
             |||||||||||| ||||| ||||||||||||||||||||||||||||||||||||||||||
Db       2581 MAKNEVFCVDPAKGGKKPARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPA 2640

Qy       2641 QRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTE 2700
             ||||:|||||||||||||||||||||||||||||||:|||||||| ||||||||
Db       2641 QRVEYLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYQACSLPEEARTAIHSLTE 2700

Qy       2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPTMLVCGD 2760
             |||||||||||||||||||||||||||||||||||||||||||||||||:|||||||||
Db       2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIVAPTMLVCGD 2760

Qy       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPRG 2820

Qy       2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDT 2880
             |||||||||||||:|||||||||||||:|||||||||||||| ||||||||||||| |||
Db       2821 RRRYYLTRDPTTPLARAAWETVRHSPINSWLGNIIQYAPTIWVRMVLMTHFFSILMVQDT 2880

Qy       2881 LDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2940
             ||||||||||:||||:||||||||||||||||:|||: |||||||||||||||||||||
Db       2881 LDQNLNFEMYGSVYSVNPLDLPAIIERLHGLDAFSMHTYSHHELTRVASALRKLGAPPLR 2940

Qy       2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000
              |||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
Db       2941 VWKSRARAVRASLISRGGKAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000

Qy       3001 DIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
              ||:||||||||| ||||||||||||||||||||
Db       3001 DIFHSVSRARPRSLLFGLLLLFVGVGLFLLPAR 3033
```

```
RESULT 7
AAR33539
ID   AAR33539 standard; protein; 3033 AA.
XX
AC   AAR33539;
XX
DT   27-AUG-2003  (revised)
DT   25-MAR-2003  (revised)
DT   01-JUL-1993  (first entry)
XX
DE   NANBH virus strain HC-J8 protein.
XX
KW   Non A non B hepatitis virus; plasma; degenerate.
XX
OS   Non-A.
OS   non-B hepatitis virus.
XX
PN   EP532167-A2.
XX
PD   17-MAR-1993.
XX
PF   30-JUL-1992;   92EP-00306952.
XX
PR   09-AUG-1991;   91JP-00287402.
PR   05-DEC-1991;   91JP-00360441.
XX
PA   (IMMO ) IMMUNO JAPAN INC.
XX
PI   Okamoto H,  Nakamura T;
XX
DR   WPI; 1993-087166/11.
DR   P-PSDB; AAQ38221.
XX
PT   Polynucleotide(s), polypeptide(s) and antibodies of NANBH virus - useful
PT   for detecting NANBH, as a vaccine and for screening blood samples.
XX
PS   Claim 8; Page 65-79; 93pp; English.
XX
CC   RNA was isolated from the plasma of human patients positive for NANBH
CC   virus (strain HC-J8) and was subjected to reverse transcription to
CC   produce cDNA. The resulting cDNA was amplified by PCR, and nucleic acid
CC   sequences determined by analysis of clones obtd. by PCR amplification (42
CC   clones in total). The NANBH HC-J8 genome was found to contain a
CC   degenerate open reading frame encoding polypeptide precursors of 3033
CC   amino acid residues. See also AAR33538 and AAR33214. (Updated on 25-MAR-
CC   2003 to correct PN field.) (Updated on 27-AUG-2003 to correct OS field.)
XX
SQ   Sequence 3033 AA;
```

Query Match 86.5%; Score 13944; DB 2; Length 3033;
Best Local Similarity 83.7%; Pred. No. 0;
Matches 2540; Conservative 247; Mismatches 246; Indels 0; Gaps 0;

```
Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSTNPKPQRKYKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              |||||||||||||||||||||||||||||| |||||||||||||:||| ||||||||:|
Db         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGCGWAGWLLSPRGSRPTWGPTDPRHRSRNLG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
              :||||:|||||||||||||||:|||||||||||||||:|:|||||||||||||||||||
Db        121 RVIDTITCGFADLMGYIPVVGAPVGGVARALAHGVRVLEDGINYATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
              ||||:| |||| ||:|||: |  ||||:|:|||||  ||||:|||||||        |||
Db        181 LLSCVTVPVSAVEVRNISSSYYATNDCSNNSITWQLTDAVLHLPGCVPCENDNGTLHCWI 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
              .|:|||||:  ||||: ||:|:||:|||:|| ||||||||:|| ||: :| |:|||| | |
Db        241 QVTPNVAVKHRGALTRSLRSHVDMIVMAATACSALYVGDVCGAVMILSQAFMVSPQRHNF 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
              |:|||||| | ||||||||||::|||| ||||||| ||||:::|| | ||||:||||
Db        301 TQECNCSIYQGHITGHRMAWDMMLSWSPTLTMILAYAARVPELVLEIIFGGHWGVVFGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
              ||||||||||: |||| ||||| |:: |  | :|      ||  | :| : |:||||||
Db        361 YFSMQGAWAKVIAILLLVAGVDATTYSSGQEAGRTVAGFAGLFTTGAKQNLYLINTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
              ||||:|||||||| |||:||||||| |||||||||:|:|| :: ||:||| |:|| ||||
Db        421 HINRSALNCNDSLQTGFLASLFYTHKFNSSGCPERLSSCRGLDDFRIGWGTLEYETNVTN 480

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
              |||||||||||| ||:| |:|||||||||||||||||: | ||||||||||||||
Db        481 DGDMRPYCWHYPPRPCGIVPARTVCGPVYCFTPSPVVVGTTDKQGVPTYTWGENETDVFL 540

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
              ||||||| |:|||||||| :|:|||||||||| | |:|:: |||||||||||||| ||||
Db        541 LNSTRPPRGAWFGCTWMNGTGFTKTCGAPPCRIRKDYNSTIDLLCPTDCFRKHPDATYLK 600

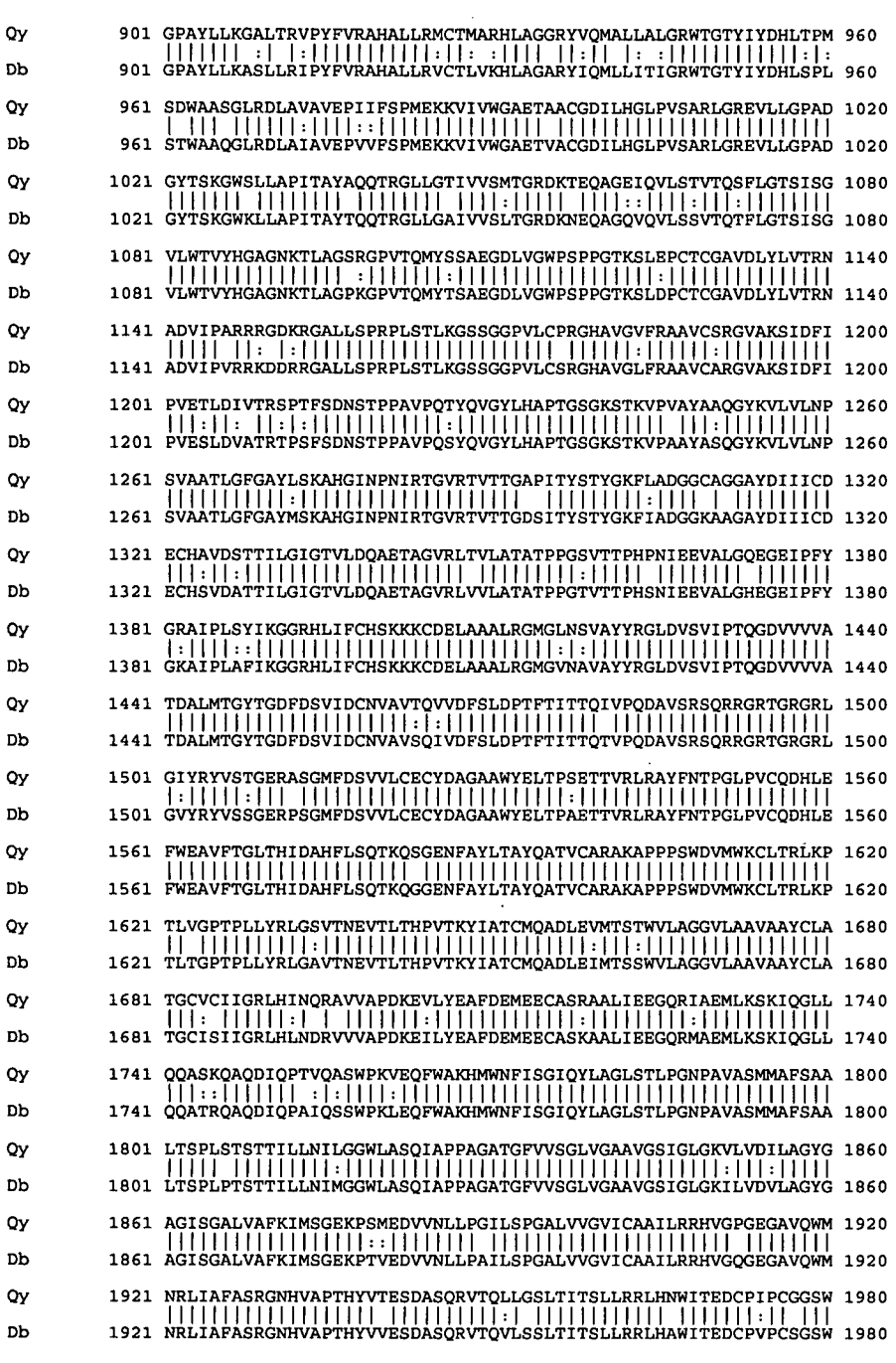
Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
              ||:|||||||||:|||||||||||||:|||| |||||||||| :|||||||||| |||
Db        601 CGAGPWLTPRCLVDYPYRLWHYPCTVNFTIFKARMYVGGVEHRFSAACNFTRGDRCRLED 660

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
              ||| | |||||||||||:||||:|||||||||||||||||||::|||||||:|||:|
Db        661 RDRGQQSPLLHSTTEWAVLPCSFSDLPALSTGLLHLHQNIVDVQYLYGLSPALTRYIVKW 720

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
              |||||||||||||:||||||||:||||||||||:|||:|||| || |:| ||| |||
Db        721 EWVILLFLLLADARICACLWMLIILGQAEAALEKLIILHSASAASANGPLWFFIFFTAAW 780

Qy        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
              |:||||||:||||: ||||| ||:|||||||||| ||:  |::| |:||:|::||||| ||
Db        781 YLKGRVVPVATYSVLGLWSFLLLVLALPQQAYALDAAEQGELGLAILVIISIFTLTPAYK 840

Qy        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
              |||| :||| |:| | || :|:| ||::|||||||||   | :| :||::|||||:|
Db        841 ILLSRSVWWLSYMLVLAEAQIQQWVPPLEVRGGRDGIIWVAVILHPRLVFEVTKWLLAIL 900
```


Qy 901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
Db 901 GPAYLLKASLLRIPYFVRAHALLRVCTLVKHLAGARYIQMLLITIGRWTGTYIYDHLSPL 960
Qy 961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
Db 961 STWAAQGLRDLAIAVEPVVFSPMEKKVIVWGAETVACGDILHGLPVSARLGREVLLGPAD 1020
Qy 1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080
Db 1021 GYTSKGWKLLAPITAYTQQTRGLLGAIVVSLTGRDKNEQAGQVQVLSSVTQTFLGTSISG 1080
Qy 1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
Db 1081 VLWTVYHGAGNKTLAGPKGPVTQMYTSAEGDLVGWPSPPGTKSLDPCTCGAVDLYLVTRN 1140
Qy 1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200
Db 1141 ADVIPVRRKDDRRGALLSPRPLSTLKGSSGGPVLCSRGHAVGLFRAAVCARGVAKSIDFI 1200
Qy 1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260
Db 1201 PVESLDVATRTPSFSDNSTPPAVPQSYQVGYLHAPTGSGKSTKVPAAYASQGYKVLVLNP 1260
Qy 1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320
Db 1261 SVAATLGFGAYMSKAHGINPNIRTGVRTVTTGDSITYSTYGKFIADGGKAAGAYDIIICD 1320
Qy 1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380
Db 1321 ECHSVDATTILGIGTVLDQAETAGVRLVVLATATPPGTVTTPHSNIEEVALGHEGEIPFY 1380
Qy 1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440
Db 1381 GKAIPLAFIKGGRHLIFCHSKKKCDELAAALRGMGVNAVAYYRGLDVSVIPTQGDVVVVA 1440
Qy 1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500
Db 1441 TDALMTGYTGDFDSVIDCNVAVSQIVDFSLDPTFTITTQTVPQDAVSRSQRRGRTGRGRL 1500
Qy 1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560
Db 1501 GVYRYVSSGERPSGMFDSVVLCECYDAGAAWYELTPAETTVRLRAYFNTPGLPVCQDHLE 1560
Qy 1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
Db 1561 FWEAVFTGLTHIDAHFLSQTKQGGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
Qy 1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680
Db 1621 TLTGPTPLLYRLGAVTNEVTLTHPVTKYIATCMQADLEIMTSSWVLAGGVLAAVAAYCLA 1680
Qy 1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740
Db 1681 TGCISIIGRLHLNDRVVVAPDKEILYEAFDEMEECASKAALIEEGQRMAEMLKSKIQGLL 1740
Qy 1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
Db 1741 QQATRQAQDIQPAIQSSWPKLEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
Qy 1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860
Db 1801 LTSPLPTSTTILLNIMGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKILVDVLAGYG 1860
Qy 1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920
Db 1861 AGISGALVAFKIMSGEKPTVEDVVNLLPAILSPGALVVGVICAAILRRHVGQGEGAVQWM 1920
Qy 1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980
Db 1921 NRLIAFASRGNHVAPTHYVVESDASQRVTQVLSSLTITSLLRRLHAWITEDCPVPCSGSW 1980
Qy 1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040
Db 1981 LQDIWDWVCSILTDFKNWLSSKLLPKMPGIPFISCQKGYKGVWAGTGVMTTRCPCGANIS 2040
Qy 2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100
Db 2041 GHVRMGTMKITGPKTCLNLWQGTFPINCYTEGPCVPKPPPNYKTAIWRVAASEYVEVTQH 2100
Qy 2101 GSYHYITGLTTDNLKVPCQLPSPEPFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160
Db 2101 GSFSYVTGLTSDNLKVPCQVPAPEPFSWVDGVQIHRFAPVPGPFFRDEVTFTVGLNSFVV 2160
Qy 2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220
Db 2161 GSQLPCDPEPDTEVLASMLTDPSHITAEAAARRLARGSPPSQASSSASQLSAPSLKATCT 2220
Qy 2221 THGKAYDVDMVDANLFMGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYMLPKKR 2280
Db 2221 THKTAYDCDMVDANLFMGGDVTRIESDSKVIVLDSLDSMTEVEDDREPSVPSEYLIKRRK 2280
Qy 2281 FPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLSEDSI 2340
Db 2281 FPPALPPWARPDYNPVLIETWKRPGYEPPTVLGCALPPTPQTPVPPPRRRRAKVLTQDNV 2340
Qy 2341 GDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTPPDELALSETGSISSMPPLEGELG 2400
Db 2341 EGVLREMADKVLSPLQDNNDSGHSTGADTGGDIVQQPSDETAASEAGSLSSMPPLEGEPG 2400
Qy 2401 DPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460
Db 2401 DPDLEFEPVGSAPPSEGECEVIDSDSKSWSTVSDQEDSVICCSMSYSWTGALITPCGPEE 2460
Qy 2461 EKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIKLAASKV 2520
Db 2461 EKLPINPLSNSLMRFHNKVYSTTSRSASLRAKKVTFDRVQVLDAHYDSVLQDVKRAASKV 2520
Qy 2521 TARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSETPIPTTI 2580
Db 2521 SARLLTVEEACALTPPHSAKSRYGFGAKEVRSLSRRAVNHIRSVWEDLLEDEHTPIDTTI 2580
Qy 2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPA 2640
Db 2581 MAKNEVFCIDPTKGGKKPARLIVYPDLGVRVCEKMALYDIAQKLPKAIMGPSYGFQYSPA 2640
Qy 2641 QRVEPLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTE 2700
Db 2641 ERVDFLLKAWGSKKDPMGFSYDTRCFDSTVTERDIRTEESIYQACSLPQEARTVIHSLTE 2700
Qy 2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPTMLVCGD 2760
Db 2701 RLYVGGPMTNSKGQSCGYRRCRASGVFTTSMGNTMTCYIKALAACKAAGIVDPVMLVCGD 2760
Qy 2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820
Db 2761 DLVVISESQGNEEDERNLRAFTEAMTRYSAPPGDLPRPEYDLELITSCSSNVSVALDSRG 2820
Qy 2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDT 2880
Db 2821 RRRYFLTRDPTTPITRAAWETVRHSPVNSWLGNIIQYAPTIWVRMVIMTHFFSILLAQDT 2880
Qy 2881 LDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2940
Db 2881 LNQNLNFEMYGAVYSVNPLDLPAIIERLHGLEAFSLHTYSPHELSRVAATLRKLGAPPLR 2940
Qy 2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000
Db 2941 AWKSRARAVRASLIAQGARAAICGRYLFNWAVKTKLKLTPLPEASRLDLSGWFTVGAGGG 3000
Qy 3001 DIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
Db 3001 DIYHSVSHARPRLLLLCLLLLSVGVGIFLLPAR 3033

RESULT 8
AAR33214
ID AAR33214 standard; protein; 3033 AA.
XX
AC AAR33214;
XX
DT 27-AUG-2003 (revised)
DT 25-MAR-2003 (revised)
DT 01-JUL-1993 (first entry)
XX
DE NANBH virus strain HC-J8 protein.
XX
KW Non A non B hepatitis virus; plasma; degenerate.
XX
OS Non-A.
OS non-B hepatitis virus.
XX
PN EP532167-A2.
XX
PD 17-MAR-1993.
XX
PF 30-JUL-1992; 92EP-00306952.
XX
PR 09-AUG-1991; 91JP-00287402.
PR 05-DEC-1991; 91JP-00360441.
XX
PA (IMMO) IMMUNO JAPAN INC.
XX
PI Okamoto H, Nakamura T;
XX
DR WPI; 1993-087166/11.
DR P-PSDB; AAQ38221.
XX
PT Polynucleotide(s), polypeptide(s) and antibodies of NANBH virus - useful
PT for detecting NANBH, as a vaccine and for screening blood samples.
XX
PS Claim 9; Page 80-93; 93pp; English.
XX
CC RNA was isolated from the plasma of human patients positive for NANBH
CC virus (strain HC-J8) and was subjected to reverse transcription to
CC produce cDNA. The resulting cDNA was amplified by PCR, and nucleic acid
CC sequences determined by analysis of clones obtd. by PCR amplification (42
CC clones in total). The NANBH HC-J8 genome was found to contain a
CC degenerate open reading frame encoding polypeptide precursors of 3033
CC amino acid residues. See also AAR33538-9. (Updated on 25-MAR-2003 to
CC correct PN field.) (Updated on 27-AUG-2003 to correct OS field.)
XX
SQ Sequence 3033 AA;

Query Match 85.6%; Score 13805; DB 2; Length 3033;
Best Local Similarity 83.3%; Pred. No. 0;
Matches 2525; Conservative 246; Mismatches 262; Indels 0; Gaps 0;

```
Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              |||||||||||||||||||||||||||| |||||||||||||:||| ||||||||:|
Db         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGCGWAGWLLSPRGSRPTWGPTDPRHRSRNLG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
              :||||:|||||||||||||||:|||||||||||||||||:|:||||||||||||||||||
Db        121 RVIDTITCGFADLMGYIPVVGAPVGGVARALAHGVRVLEDGINYATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
              ||||:| |||| ||:|||: | ||||:|:|||||  ||||:|||||||      :|||
Db        181 LLSCVTMPVSAVEVRNISSSYYATNDCSNNSITWQLTDAVLHLPGCVPCENDNGTLRCWI 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
                |:|:|||:  |||| |||||:||:||:| ||||||||:|| ||: :| |:|||| | |
Db        241 QVTPDVAVKHRGALTRSLRTHVDMIVMAATACSALYVGDVCGAVMILSQAFMVSPQRHNF 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
              |:|||||| | |||||||||||::|||| |||||| ||||:::||| | |||| |||
Db        301 TQECNCSIYQGHITGHRMAWDMMLDWSPTLAMILAYAARVPELVLEIIFGGHWGVAFGLG 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
              |||||||||||| |||| ||||| |:: |  | :    ::|||: | :| : |:||||||
Db        361 YFSMQGAWAKVVAILLLVAGVDASTYSTGQQAGRAAYGISSLFNTGAKQNLHLINTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
              |||||||||||| |||||| |  |||||||||:|:|| :: ||:||| |:|| ||||
Db        421 HINRTALNCNDSLETGFIASLVYYRRFNSSGCPERLSSCRGLDDFRIGWGTLEYETNVTN 480

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
              |||||||||||| ||:| |:|||||||||||||||||||: | ||||||||||||||
Db        481 DEDMRPYCWHYPPRPCGIVPARTVCGPVYCFTPSPVVVGTTDKQGVPTYTWGENETDVFL 540

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
              ||||||| |:||| |||| :|:||| |||||| | |:|:: |||||||||||||| ||||
Db        541 LNSTRPPRGAWFGKTWMNGTGFTKTKGAPPCRIRKDYNSTIDLLCPTDCFRKHPDATYLK 600

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
              ||:|||||||:|||||||||||||:|||| |||||||||| :|| |||||||| |||
Db        601 CGAGPWLTPRCLVDYPYRLWHYPCTVNFTIFKARMYVGGVEHRFSAAKNFTRGDRCRLED 660

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
              ||| | |||||||||||: |||:||||||||||||||||||::||||||||:|||:|
Db        661 RDRGQQSPLLHSTTEWAVFPCSFSDLPALSTGLLHLHQNIVDVQYLYGLSPALTRYIVKW 720

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
              |||||||||||||||||||| :|||||||||||:|||:||||| || |:| ||| |||
Db        721 EWVILLFLLLADARVCACLWMLNILGQAEAALEKLIILHSASAASANGPLWFFIFFTAAW 780

Qy        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
              |:|||||:|||: ||||| ||:|||||||||  |::| |:||:|::||||| ||
Db        781 YLKGRVVPVATYSVLGLWSFLLLVLALPQQAYALDAAEQGELGLAILVIISIFTLTPAYK 840

Qy        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
              |||| :||| |:| | || :|:| ||::|||||||||   | :| :||::|||||:|
Db        841 ILLSRSVWWLSYMLVLAEAQIQQWVPPLEVRGGRDGIIWVAVILHPRLVFEVTKWLLAIL 900

Qy        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
              ||||||: :| |:|||||||||:||: :|||| |:|| |: :|||||||||||:|:
Db        901 GPAYLLRASLLRIPYFVRAHALLRVCTLVKHLAGASYIQMLLITIGRWTGTYIYDHLSPL 960

Qy        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
              | ||| |||||:|||::|||||||||||||| ||||||||||||||||||||||||||
Db        961 STWAAQGLRDLAIAVEPVVFSPMEKKVIVWGAETVACGDILHGLPVSARLGREVLLGPAD 1020

Qy       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080
              |||||||:|||||||| |||||||| ||||:|||| ||||::||||:|||:||||||||
Db       1021 GYTSKGWNLLAPITAYTQQTRGLLGAIVVSLTGRDKNEQAGQVQVLSSVTQTFLGTSISG 1080

Qy       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
              |||||||||||||||| :|||||||:|||||||||||||||||:|||||||||||||||
Db       1081 VLWTVYHGAGNKTLAGPKGPVTQMYTSAEGDLVGWPSPPGTKSLDPCTCGAVDLYLVTRN 1140

Qy       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200
              ||||| ||: |:|||||||||||||||||||| ||||||:||||||:|||||||||||
Db       1141 ADVIPVRRKDDRRGALLSPRPLSTLKGSSGGPVLCSRGHAVGLFRAAVCARGVAKSIDFI 1200

Qy       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260
              |||:||| ||:|:||||| ||||||:|||||||||||||||||| |||:||||||||||
Db       1201 PVESLDIATRTPSFSDNSAPPAVPQSYQVGYLHAPTGSGKSTKVPAAYASQGYKVLVLNP 1260

Qy       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320
              ||||||||||:|||||||||||||||||  |||||||||:|||||| |||||||||
Db       1261 SVAATLGFGAYMSKAHGINPNIRTGVRTVTTGDSITYSTYGKFIADGGCAAGAYDIIICD 1320

Qy       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380
              |||:||:|||||||||||||||||||| |||||||||:||||| |||||||| |||||||
Db       1321 ECHSVDATTILGIGTVLDQAETAGVRLVVLATATPPGTVTTPHSNIEEVALGHEGEIPFY 1380

Qy       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440
```

```
Db       1381 GKAIPLAFIKGGRHLIFCHSKKKCDELAAALRGTGVNAVAYYRGLDVSVIPTQGDVVVVA 1440

Qy       1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500

Db       1441 TDALMTGYTGDFDSVIDCNVAVSQIVDFSLDPTFTITTQTVPQDAVSRSQRRGRTGRGRL 1500

Qy       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560

Db       1501 GIYRYVSSGEGPSGMFDSVVPCECYDAGAAWYELTPAETTVRLRAYFNTPGLPVCQDHLE 1560

Qy       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620

Db       1561 FWEAVFTGLTHINAHFLSQTKQGGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620

Qy       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680

Db       1621 TLTGPTPLLYRLGAVTNEVTLTHPVYKYIATCMQADLEIMTSSWVLAGGVLAAVAAYCLA 1680

Qy       1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740

Db       1681 TGCISIIGRLHLNDRVVVTPDKEILYEAFDEMEECASKAALIEEGQRMAEMLKSKIQGLL 1740

Qy       1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800

Db       1741 QQATRQAQGMQPAIQSSWPKLEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800

Qy       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860

Db       1801 LTSPLPTSTTILLNIMGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKILVDVLAGYG 1860

Qy       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920

Db       1861 AGISGALVAFKIMSGEKPTVEDVVNLLPAILSPGALVVGVICAAILRRHVGQGEGAVQWM 1920

Qy       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980

Db       1921 NRLIAFASRGNHVAPTHYVVESDASQRVTQVLSSLTITSLLRRLHAWITEDCPIPCSGSW 1980

Qy       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040

Db       1981 LQDIWDWVCSILTDFKNWLSSKLLPKMPGIPFISCQKGYKGVWAGTGVMTTRYPCGANIS 2040

Qy       2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100

Db       2041 GHVRMGTMKITGPKTCLNLWQGTFPINCYTEGPCVPKPPPNYKTAIWRVAASEYVEVTQH 2100

Qy       2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160

Db       2101 GSFSYVTGLTSDNLKVPCQVPAPEFFSWVDGVQIHRFAPVPGPFFRDEVTFTVGLNSFVV 2160

Qy       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220

Db       2161 GSQLPCDPEPDTEVLASMLTDPSHITAEAAARRLARGSPPSQASSSASQLSAPSLKATCT 2220

Qy       2221 THGKAYDVDMVDANLFMGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYMLPKKR 2280

Db       2221 THKTAYDCDMVDANLFMGGDVTRIESDSKVIVLDSLDSMTEVEDDREPSVPSEYLIKRRK 2280

Qy       2281 FPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLSEDSI 2340

Db       2281 FPPALPPWARPDYNPVLIETWKRPGYEPPTVLGCALPPTLQTPVPPPRRRRAKILTQDDV 2340

Qy       2341 GDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTPPDELALSETGSISSMPPLEGELG 2400

Db       2341 EGILREMADKVLSPLQDNNDSGHSTGADTGGDIVQQPSDETAASEAGSLSSMPPLEGEPG 2400

Qy       2401 DPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460

Db       2401 DPDLEFEPVGSAPPSEGECEVIDSDSKSWSTVSDQEDSVICCSMSYSWTGALITPCGPEE 2460

Qy       2461 EKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIKLAASKV 2520

Db       2461 EKLPINPLSNSLMRFHNKVYSTTSRSASLRAKKVTFDRVQVLDAHYDSVLQDVKRAASKV 2520

Qy       2521 TARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSETPIPTTI 2580

Db       2521 GARLLTVEEACALTPPHSAKSRYGFGAKEVRSLSRRAVNHIRSVWENLLEDQRTPIDTTI 2580

Qy       2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPA 2640

Db       2581 MAKNEVFCIDPTKGGKKPARLIVYPDLGVRVCEKMALYDITQKLPKAIMGPSYGFQYSPA 2640

Qy       2641 QRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTE 2700

Db       2641 ERVDFLLKAWGSKKDPMGFSYDTRKFDSTVTERDIRTEESIYQACSLPQEARTVIHSLTE 2700

Qy       2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPTMLVCGD 2760

Db       2701 RLYVGGPMTNSKGQSCGYRRCRISGVFTTSMGNTMTCYIKALAACKAAGIVDPVMLVCGD 2760

Qy       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820

Db       2761 DLVVISESQGNEEDERNLRIFTEAMTRYSAPPGDLPRPEYDLELITSCSSNVSVALDSRG 2820

Qy       2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDT 2880

Db       2821 RRRYFLTRDPTTPITRAAWETVRHSPVNSWLGNIIQYAPTIWVRMVIMTHFFSILLAQDT 2880

Qy       2881 LDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2940

Db       2881 LNQNLNFEMYGAVYSVNPLDLPAIIERLHGLEAFSLHTYSPHELSRVAATLRKLGAPPLR 2940

Qy       2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000

Db       2941 AWKSRARAVRASLIAQGARAAICGRYLFNWAVKTKLKLTPLPEASRLDLSGWFTVGAGGG 3000

Qy       3001 DIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033

Db       3001 DIYHSVSHARPRLLLLCLLLLSVGVGIFLLPAR 3033
```

RESULT 9
AAB30730
ID AAB30730 standard; protein; 3015 AA.
XX
AC AAB30730;
XX
DT 02-APR-2001 (first entry)
XX
DE Amino acid sequence of chimeric Hepatitis C virus clone pH77CV-J6S.
XX
KW HCV; HCV strain HC-J6CH; HCV genotype 2a; antiviral; vaccine.
XX
OS Synthetic.
OS Hepatitis C virus.
XX
PN WO200075338-A2.
XX
PD 14-DEC-2000.
XX
PF 02-JUN-2000; 2000WO-US015446.
XX
PR 04-JUN-1999; 99US-0137693P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Yanagi M, Bukh J, Emerson SU, Purcell RH;
XX
DR WPI; 2001-061728/07.
DR N-PSDB; AAC86645.
XX
PT Nucleic acid molecule encoding human hepatitis C virus of genotype 2a for
PT developing vaccines, for diagnosis of hepatitis C virus and in screening
PT assays for identification of antiviral agents.
XX

PS Disclosure; Page 103-115; 167pp; English.
XX
CC AAB30730-33 are encoded by chimeric cDNA clones of infectious Hepatitis C
CC virus (HCV). In each clone the C, E1 and E2 genes are derived from HCV
CC strain HC-J6CH genotype 2a. The p7 protein was derived from HCV strain HC
CC -J6CH or HCV strain pCV-H77C, and the NS genes were derived from HCV
CC strain pCV-H77C. Such HCV sequences are capable of expressing the virus
CC when transfected into cells. The HCV protein is useful for assaying
CC candidate antiviral agents for activity against HCV. Antibodies specific
CC for HCV polypeptide are useful in prevention and treatment of diseases
CC caused by HCV in animals, in particular humans. The HCV polypeptides
CC serve as immunogens in the development of vaccines for preventing HCV in
CC mammals or as antigens in diagnostic assays for detecting the presence of
CC HCV in biological samples. The HCV polynucleotide is also useful for
CC identifying cell lines capable of supporting the replication of HCV in
CC vitro and to produce attenuated viral strains via passage in vitro or in
CC vivo
XX
SQ Sequence 3015 AA;

Query Match 80.9%; Score 13049; DB 4; Length 3015;
Best Local Similarity 78.8%; Pred. No. 0;
Matches 2395; Conservative 261; Mismatches 353; Indels 30; Gaps 4;

```
Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780

Qy        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
              |||||||||||||||||||||||||||||| | |  |  :|| :  ||:| ||
Db        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYALDTEVAASCGGVVLVGLMALTLSPYYK 840

Qy        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
              :|  :||| | ||  || :  | ||: |||||| :|  : : :| :|||||| |||:
Db        841 RYISWCMWWLQYFLTRVEAQLHVWVPPLNVRGGRDAVILLMCVVHPTLVFDITKLLLAIF 900

Qy        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
              || ::|: :| :|||||   |||:| :|| :||| |||||:: ||  ||||:|:||||:
Db        901 GPLWILQASLLKVPYFVRVQGLLRICALARKIAGGHYVQMAIIKLGALTGTYVYNHLTPL 960

Qy        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
              ||| :|||||||||||::|| || |:| |||:|||||||::||||||| |:|:||||||
Db        961 RDWAHNGLRDLAVAVEPVVFSRMETKLITWGADTAACGDIINGLPVSARRGQEILLGPAD 1020

Qy       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080
              |  |||| |||||||||||||||||| |: |:||||| :  ||:|::|| ||:|| | |:|
Db       1021 GMVSKGWRLLAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTATQTFLATCING 1080

Qy       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
              | |||||||| :|:|  :||| ||:: : ||||||:|  |::|| ||||| |||||:
Db       1081 VCWTVYHGAGTRTIASPKGPVIQMYTNVDQDLVGWPAPQGSRSLTPCTCGSSDLYLVTRH 1140

Qy       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200
              ||||| ||||| ||:|||||:| |||||||:||| |||||:||||||:|||||::|||
Db       1141 ADVIPVRRRGDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGLFRAAVCTRGVAKAVDFI 1200

Qy       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260
              ||| |   ||| |:|||:||||||::|| :|||||||||||||| |||||||||||||
Db       1201 PVENLGTTMRSPVFTDNSSPPAVPQSFQVAHLHAPTGSGKSTKVPAAYAAQGYKVLVLNP 1260

Qy       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320
              |||||||||:|||||::|||||||:|||:|||||||||||||:|||||||||
Db       1261 SVAATLGFGAYMSKAHGVDPNIRTGVRTITTGSPITYSTYGKFLADGGCSGGAYDIIICD 1320

Qy       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380
              |||: |:|:|||||||||||||| || |||||||||||  |||||||  ||||||
Db       1321 ECHSTDATSILGIGTVLDQAETAGARLVVLATATPPGSVTVSHPNIEEVALSTTGEIPFY 1380

Qy       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440
              |:|||  |||||||||||||||||| |  :|:|:|||||||||||| ||||||:
Db       1381 GKAIPLEVIKGGRHLIFCHSKKKCDELAAKLVALGINAVAYYRGLDVSVIPTSGDVVVVS 1440

Qy       1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500
              |||||||:|||||||||  ||| |||||||||| |  :|||||:|||||||:
Db       1441 TDALMTGFTGDFDSVIDCNTCVTQTVDFSLDPTFTIETTTLPQDAVSRTQRRGRTGRGKP 1500

Qy       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560
              ||||:|: ||| |||||| |||||||| |||||||:||||||| ||||||||||||
Db       1501 GIYRFVAPGERPSGMFDSSVLCECYDAGCAWYELTPAETTVRLRAYMNTPGLPVCQDHLE 1560

Qy       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
              ||| |||||||||||||||||||| || |||||||||:|||||| ||||| ||||
Db       1561 FWEGVFTGLTHIDAHFLSQTKQSGENFPYLVAYQATVCARAQAPPPSWDQMWKCLIRLKP 1620

Qy       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680
              || ||||||||||:| |||||||:|||| ||| ||||:|||||| ||||||:|||||:
Db       1621 TLHGPTPLLYRLGAVQNEVTLTHPITKYIMTCMSADLEVVTSTWVLVGGVLAALAAYCLS 1680

Qy       1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740
              |||| |:||: :: :  : ||:|||||: |||||||:        ||:|  :||  | |  |||
Db       1681 TGCVVIVGRIVLSGKPAIIPDREVLYQEFDEMEECSQHLPYIEQGMMLAEQFKQKALGLL 1740

Qy       1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
              | ||: |: | | || :| |:| |||||||||||||||||||||||||:||:|||:||
Db       1741 QTASRHAEVITPAVQTNWQKLEVFWAKHMWNFISGIQYLAGLSTLPGNPAIASLMAFTAA 1800

Qy       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860
              :||||:|  |:| ||||||:|:|:| |  || || :|| |||:||:||||||||||||||
```

```
Db       1801 VTSPLTTGQTLLFNILGGWVAAQLAAPGAATAFVGAGLAGAAIGSVGLGKVLVDILAGYG 1860

Qy       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920
              ||::|||||||||||| || ||:||||| ||||||||||:||||||||||||||||||||
Db       1861 AGVAGALVAFKIMSGEVPSTEDLVNLLPAILSPGALVVGVVCAAILRRHVGPGEGAVQWM 1920

Qy       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980
              |||||||||||||:||||| |||| ||| :| |||:| |||||| ||: :|  || |||
Db       1921 NRLIAFASRGNHVSPTHYVPESDAAARVTAILSSLTVTQLLRRLHQWISSECTTPCSGSW 1980

Qy       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040
              |||:|||:| :|:||| || :|| |::||:||||||:||:||| | ||| ||| ||| |:
Db       1981 LRDIWDWICEVLSDFKTWLKAKLMPQLPGIPFVSCQRGYRGVWRGDGIMHTRCHCGAEIT 2040

Qy       2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100
              |:|: |:||| ||:|| |:| |||||| || | | | ||||:| |:|||:| || |: :
Db       2041 GHVKNGTMRIVGPRTCRNMWSGTFPINAYTTGPCTPLPAPNYKFALWRVSAEEYVEIRRV 2100

Qy       2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160
              | :||::|:|||||| |||:|||||: :|||::|||||  ||  |:|||| |||: : |
Db       2101 GDFHYVSGMTTDNLKCPCQIPSPEFFTELDGVRLHRFAPPCKPLLREEVSFRVGLHEYPV 2160

Qy       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220
              ||||||:|||| || |||||||||||| | |||||||||| ||||||||||||:||||
Db       2161 GSQLPCEPEPDVAVLTSMLTDPSHITAEAAGRRLARGSPPSMASSSASQLSAPSLKATCT 2220

Qy       2221 THGKAYDVDMVDANLF----MGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYML 2276
               :  : | ::::|||     |||::||:|| :|||:||| ||:| |  : | |:|:| :
Db       2221 ANHDSPDAELIEANLLWRQEMGGNITRVESENKVVILDSFDPLVAEEDEREVSVPAEILR 2280

Qy       2277 PKKRFPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLS 2336
               :||  ||| ||||||||||||:||:|||:|  | || |||||  | ||||::||| |:
Db       2281 KSRRFARALPVWARPDYNPPLVETWKKPDYEPPVVHGCPLPPPRSPPVPPPRKKRTVVLT 2340

Qy       2337 EDSIGDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTPPDELALSETGSISSMPPLE 2396
              | ::  || :|| |||     || :| :|   :  : |  |||     |:  | |||||||
Db       2341 ESTLSTALAELATKSFGSSSTSGITGDNTTTSSEPAPSGCPPD----SDVESYSSMPPLE 2396

Qy       2397 GELGDPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSE--EDDSVVCCSMSYSWTGALIT 2454
              || |||||                      ||||| |   : : ||||||||||||||:|
Db       2397 GEPGDPDL--------------------SDGSWSTVSSGADTEDVVCCSMSYSWTGALVT 2436

Qy       2455 PCSPEEEKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIK 2514
              ||: ||:||||| |||||||:|| || ||::||  | |||||||:|||||:|  |||::|
Db       2437 PCAAEEQKLPINALSNSLLRHHNLVYSTTSRSACQRQKKVTFDRLQVLDSHYQDVLKEVK 2496

Qy       2515 LAASKVTARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSET 2574
               ||||| | ||::|||| |||||||:||:|:|||:||  : :|| || |||||||||| |
Db       2497 AAASKVKANLLSVEEACSLTPPHSAKSKFGYGAKDVRCHARKAVAHINSVWKDLLEDSVT 2556

Qy       2575 PIPTTIMAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYG 2634
              || |||||||||||| | |||:| |||||:|||||||||||||:  ||| |||||:|||
Db       2557 PIDTTIMAKNEVFCVQPEKGGRKPARLIVFPDLGVRVCEKMALYDVVSKLPLAVMGSSYG 2616

Qy       2635 FQYSPAQRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTA 2694
              ||||| |||||::||  || |||||||||||||||| ||||||:||: | |  :|  |
Db       2617 FQYSPGQRVEFLVQAWKSKKTPMGFSYDTRCFDSTVTESDIRTEEAIYQCCDLDPQARVA 2676

Qy       2695 IHSLTERLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPT 2754
              | |||||||||: ||:|: |||||||||||||| |||:|||:|| |||:|||:   |
Db       2677 IKSLTERLYVGGPLTNSRGENCGYRRCRASGVLTTSCGNTLTCYIKARAACRAAGLQDCT 2736

Qy       2755 MLVCGDDLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSV 2814
              |||||||||| || | :||  :|||||||||||||||||:|||||||||||||||||
Db       2737 MLVCGDDLVVICESAGVQEDAASLRAFTEAMTRYSAPPGDPPQPEYDLELITSCSSNVSV 2796

Qy       2815 ALGPQGRRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSI 2874
              |    |:| ||||||||||:|||||| ||:|||||||||| :|||:||||:|||||||:
Db       2797 AHDGAGKRVYYLTRDPTTPLARAAWETARHTPVNSWLGNIIMFAPTLWARMILMTHFFSV 2856

Qy       2875 LMAQDTLDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKL 2934
              |:|:| |:| || |:||| ||: ||||| ||:||||| |||||:|:| |: |||: ||||
Db       2857 LIARDQLEQALNCEIYGACYSIEPLDLPPIIQRLHGLSAFSLHSYSPGEINRVAACLRKL 2916
```

```
Qy       2935 GAPPLRAWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFT 2994
              | ||||||: |||:||| |:||||||:||:|||||||:|||||||:  |  |||| |||
Db       2917 GVPPLRAWRHRARSVRARLLSRGGRAAICGKYLFNWAVRTKLKLTPIAAAGRLDLSGWFT 2976

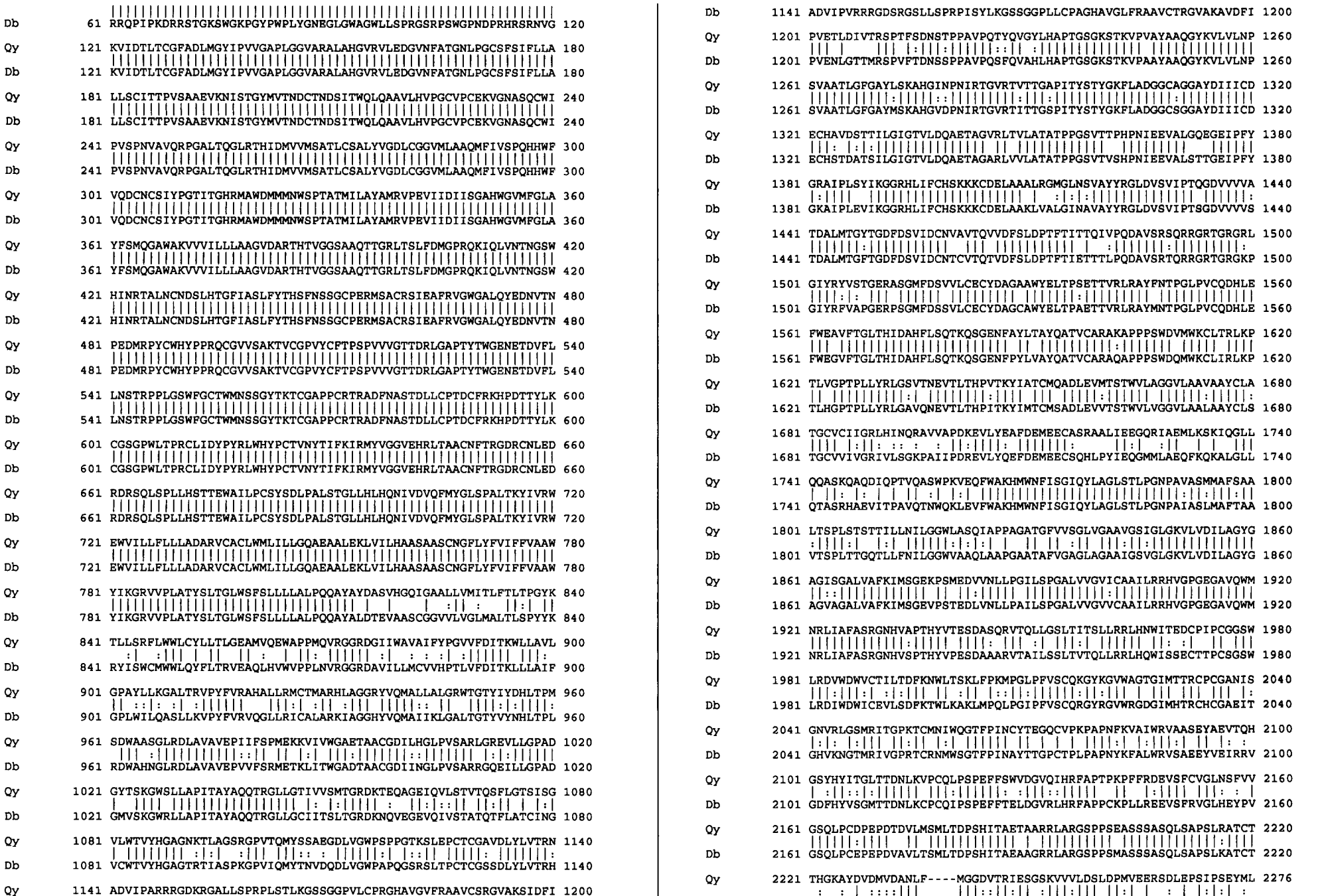
Qy       2995 VGAGGGDIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
                |  ||||||||| ||||   | ||||  |||::||| |
Db       2977 AGYSGGDIYHSVSHARPRWFWFCLLLLAAGVGIYLLPNR 3015


RESULT 10
AAB30732
ID   AAB30732 standard; protein; 3015 AA.
XX
AC   AAB30732;
XX
DT   02-APR-2001  (first entry)
XX
DE   Amino acid sequence of chimeric Hepatitis C virus clone J6S.
XX
KW   HCV; HCV strain HC-J6CH; HCV genotype 2a; antiviral; vaccine.
XX
OS   Synthetic.
OS   Hepatitis C virus.
XX
PN   WO200075338-A2.
XX
PD   14-DEC-2000.
XX
PF   02-JUN-2000; 2000WO-US015446.
XX
PR   04-JUN-1999;   99US-0137693P.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Yanagi M,  Bukh J,  Emerson SU,  Purcell RH;
XX
DR   WPI; 2001-061728/07.
DR   N-PSDB; AAC86647.
XX
PT   Nucleic acid molecule encoding human hepatitis C virus of genotype 2a for
PT   developing vaccines, for diagnosis of hepatitis C virus and in screening
PT   assays for identification of antiviral agents.
XX
PS   Disclosure; Page 134-146; 167pp; English.
XX
CC   AAB30730-33 are encoded by chimeric cDNA clones of infectious Hepatitis C
CC   virus (HCV). In each clone the C, E1 and E2 genes are derived from HCV
CC   strain HC-J6CH genotype 2a. The p7 protein was derived from HCV strain HC
CC   -J6CH or HCV strain pCV-H77C, and the NS genes were derived from HCV
CC   strain pCV-H77C. Such HCV sequences are capable of expressing the virus
CC   when transfected into cells. The HCV protein is useful for assaying
CC   candidate antiviral agents for activity against HCV. Antibodies specific
CC   for HCV polypeptide are useful in prevention and treatment of diseases
CC   caused by HCV in animals, in particular humans. The HCV polypeptides
CC   serve as immunogens in the development of vaccines for preventing HCV in
CC   mammals or as antigens in diagnostic assays for detecting the presence of
CC   HCV in biological samples. The HCV polynucleotide is also useful for
CC   identifying cell lines capable of supporting the replication of HCV in
CC   vitro and to produce attenuated viral strains via passage in vitro or in
CC   vivo
XX
SQ   Sequence 3015 AA;

  Query Match             80.9%;  Score 13049;  DB 4;  Length 3015;
  Best Local Similarity   78.8%;  Pred. No. 0;
  Matches 2395;  Conservative  261;  Mismatches  353;  Indels   30;  Gaps    4;

Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
```



```
Qy       2277 PKKRFPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLS 2336
Db       2281 KSRRFARALPVWARPDYNPPLVETWKKPDYEPPVVHGCPLPPPRSPPVPPPRKKRTVVLT 2340

Qy       2337 EDSIGDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTPPDELALSETGSISSMPPLE 2396
Db       2341 ESTLSTALAELATKSFGSSSTSGITGDNTTTSSEPAPSGCPPD----SDVESYSSMPPLE 2396

Qy       2397 GELGDPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSE--EDDSVVCCSMSYSWTGALIT 2454
Db       2397 GEPGDPDL-------------------SDGSWSTVSSGADTEDVVCCSMSYSWTGALVT 2436

Qy       2455 PCSPEEEKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIK 2514
Db       2437 PCAAEEQKLPINALSNSLLRHHNLVYSTTSRSACQRQKKVTFDRLQVLDSHYQDVLKEVK 2496

Qy       2515 LAASKVTARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSET 2574
Db       2497 AAASKVKANLLSVEEACSLTPPHSAKSKFGYGAKDVRCHARKAVAHINSVWKDLLEDSVT 2556

Qy       2575 PIPTTIMAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYG 2634
Db       2557 PIDTTIMAKNEVFCVQPEKGGRKPARLIVFPDLGVRVCEKMALYDVVSKLPLAVMGSSYG 2616

Qy       2635 FQYSPAQRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTA 2694
Db       2617 FQYSPGQRVEFLVQAWKSKKTPMGFSYDTRCFDSTVTESDIRTEEAIYQCCDLDPQARVA 2676

Qy       2695 IHSLTERLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPT 2754
Db       2677 IKSLTERLYVGGPLTNSRGENCGYRRCRASGVLTTSCGNTLTCYIKARAACRAAGLQDCT 2736

Qy       2755 MLVCGDDLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSV 2814
Db       2737 MLVCGDDLVVICESAGVQEDAASLRAFTEAMTRYSAPPGDPPQPEYDLELITSCSSNVSV 2796

Qy       2815 ALGPQGRRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSI 2874
Db       2797 AHDGAGKRVYYLTRDPTTPLARAAWETARHTPVNSWLGNIIMFAPTLWARMILMTHFFSV 2856

Qy       2875 LMAQDTLDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKL 2934
Db       2857 LIARDQLEQALNCEIYGACYSIEPLDLPPIIQRLHGLSAFSLHSYSPGEINRVAACLRKL 2916

Qy       2935 GAPPLRAWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFT 2994
Db       2917 GVPPLRAWRHRARSVRARLLSRGGRAAICGKYLFNWAVRTKLKLTPIAAAGRLDLSGWFT 2976

Qy       2995 VGAGGGDIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
Db       2977 AGYSGGDIYHSVSHARPRWFWFCLLLLAAGVGIYLLPNR 3015
```

```
RESULT 11
AAB30733
ID   AAB30733 standard; protein; 3015 AA.
XX
AC   AAB30733;
XX
DT   02-APR-2001  (first entry)
XX
DE   Amino acid sequence of chimeric Hepatitis C virus clone pH77(p7)-J6S.
XX
KW   HCV; HCV strain HC-J6CH; HCV genotype 2a; antiviral; vaccine.
XX
OS   Synthetic.
OS   Hepatitis C virus.
XX
PN   WO200075338-A2.
XX
PD   14-DEC-2000.
XX
PF   02-JUN-2000; 2000WO-US015446.
XX
PR   04-JUN-1999;   99US-0137693P.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Yanagi M,  Bukh J,  Emerson SU,  Purcell RH;
XX
DR   WPI; 2001-061728/07.
DR   N-PSDB; AAC86648.
XX
PT   Nucleic acid molecule encoding human hepatitis C virus of genotype 2a for
PT   developing vaccines, for diagnosis of hepatitis C virus and in screening
PT   assays for identification of antiviral agents.
XX
PS   Disclosure; Page 150-161; 167pp; English.
XX
CC   AAB30730-33 are encoded by chimeric cDNA clones of infectious Hepatitis C
CC   virus (HCV). In each clone the C, E1 and E2 genes are derived from HCV
CC   strain HC-J6CH genotype 2a. The p7 protein was derived from HCV strain HC
CC   -J6CH or HCV strain pCV-H77C, and the NS genes were derived from HCV
CC   strain pCV-H77C. Such HCV sequences are capable of expressing the virus
CC   when transfected into cells. The HCV protein is useful for assaying
CC   candidate antiviral agents for activity against HCV. Antibodies specific
CC   for HCV polypeptide are useful in prevention and treatment of diseases
CC   caused by HCV in animals, in particular humans. The HCV polypeptides
CC   serve as immunogens in the development of vaccines for preventing HCV in
CC   mammals or as antigens in diagnostic assays for detecting the presence of
CC   HCV in biological samples. The HCV polynucleotide is also useful for
CC   identifying cell lines capable of supporting the replication of HCV in
CC   vitro and to produce attenuated viral strains via passage in vitro or in
CC   vivo
XX
SQ   Sequence 3015 AA;
```

Query Match 80.2%; Score 12925; DB 4; Length 3015;
Best Local Similarity 78.0%; Pred. No. 0;
Matches 2371; Conservative 270; Mismatches 368; Indels 30; Gaps 4;

```
Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
Db         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
Db        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
Db        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
Db        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
Db        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
Db        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
Db        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
```



```
Qy       2695 IHSLTERLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPT 2754
             | ||||||||||||: ||:|: |||||||||||||||| |||:|||:|| |||:|||:   |
Db       2677 IKSLTERLYVGGPLTNSRGENCGYRRCRASGVLTTSCGNTLTCYIKARAACRAAGLQDCT 2736

Qy       2755 MLVCGDDLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSV 2814
             |||||||||| || | :||  :|||||||||||||||||:||||||||||||||||
Db       2737 MLVCGDDLVVICESAGVQEDAASLRAFTEAMTRYSAPPGDPPQPEYDLELITSCSSNVSV 2796

Qy       2815 ALGPQGRRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSI 2874
             |    |:| ||||||||||:||||||| ||:|||||||||| :|||:||||:|||||||:
Db       2797 AHDGAGKRVYYLTRDPTTPLARAAWETARHTPVNSWLGNIIMFAPTLWARMILMTHFFSV 2856

Qy       2875 LMAQDTLDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKL 2934
             |:|:| |:| || |:||| ||: ||||| ||:||||| |||||:|:| |: |||: ||||
Db       2857 LIARDQLEQALNCEIYGACYSIEPLDLPPIIQRLHGLSAFSLHSYSPGEINRVAACLRKL 2916

Qy       2935 GAPPLRAWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFT 2994
             | ||||||: |||:||| |:|||||||:||:|||||||:|||||||:  |  |||| |||
Db       2917 GVPPLRAWRHRARSVRARLLSRGGRAAICGKYLFNWAVRTKLKLTPIAAAGRLDLSGWFT 2976

Qy       2995 VGAGGGDIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
              |  ||||||||| ||||   | ||||  |||::||| |
Db       2977 AGYSGGDIYHSVSHARPRWFWFCLLLLAAGVGIYLLPNR 3015
```

RESULT 12
AAB30731
ID AAB30731 standard; protein; 3015 AA.
XX
AC AAB30731;
XX
DT 02-APR-2001 (first entry)
XX
DE Amino acid sequence of chimeric Hepatitis C virus clone pH77(p7)CV-J6S.
XX
KW HCV; HCV strain HC-J6CH; HCV genotype 2a; antiviral; vaccine.
XX
OS Synthetic.
OS Hepatitis C virus.
XX
PN WO200075338-A2.
XX
PD 14-DEC-2000.
XX
PF 02-JUN-2000; 2000WO-US015446.
XX
PR 04-JUN-1999; 99US-0137693P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Yanagi M, Bukh J, Emerson SU, Purcell RH;
XX
DR WPI; 2001-061728/07.
DR N-PSDB; AAC86646.
XX
PT Nucleic acid molecule encoding human hepatitis C virus of genotype 2a for
PT developing vaccines, for diagnosis of hepatitis C virus and in screening
PT assays for identification of antiviral agents.
XX
PS Disclosure; Page 119-130; 167pp; English.
XX
CC AAB30730-33 are encoded by chimeric cDNA clones of infectious Hepatitis C
CC virus (HCV). In each clone the C, E1 and E2 genes are derived from HCV
CC strain HC-J6CH genotype 2a. The p7 protein was derived from HCV strain HC
CC -J6CH or HCV strain pCV-H77C, and the NS genes were derived from HCV
CC strain pCV-H77C. Such HCV sequences are capable of expressing the virus
CC when transfected into cells. The HCV protein is useful for assaying
CC candidate antiviral agents for activity against HCV. Antibodies specific
CC for HCV polypeptide are useful in prevention and treatment of diseases
CC caused by HCV in animals, in particular humans. The HCV polypeptides
CC serve as immunogens in the development of vaccines for preventing HCV in
CC mammals or as antigens in diagnostic assays for detecting the presence of
CC HCV in biological samples. The HCV polynucleotide is also useful for
CC identifying cell lines capable of supporting the replication of HCV in
CC vitro and to produce attenuated viral strains via passage in vitro or in
CC vivo
XX
SQ Sequence 3015 AA;

Query Match 80.2%; Score 12925; DB 4; Length 3015;
Best Local Similarity 78.0%; Pred. No. 0;
Matches 2371; Conservative 270; Mismatches 368; Indels 30; Gaps 4;

```
Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660

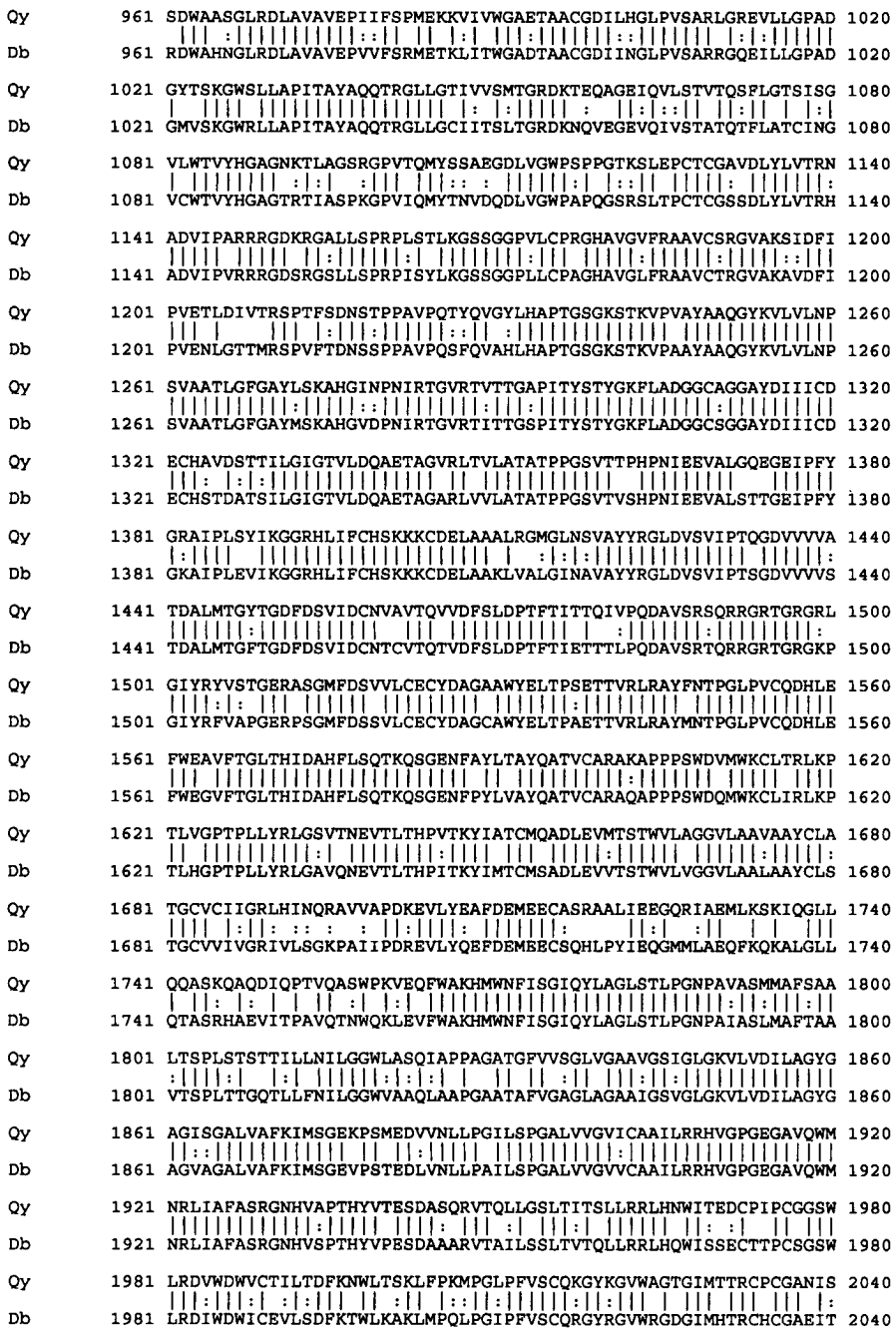
Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
              |||||||||||||||||||||||||||||| ||||:||| |  :| : |::||  ||
Db        721 EWVILLFLLLADARVCACLWMLILLGQAEAALENLVILNAASLAGTHGLVSFLVFFCFAW 780

Qy        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
              |:||| || | |:| |:|   ||||||||:||| |  |     |  :|| :   ||:| ||
Db        781 YLKGRWVPGAVYALYGMWPLLLLLLALPQRAYALDTEVAASCGGVVLVGLMALTLSPYYK 840

Qy        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
              :|  :||| | ||  || :  | ||: |||||| :|  : : :| :|||||| |||:
Db        841 RYISWCMWWLQYFLTRVEAQLHVWVPPLNVRGGRDAVILLMCVVHPTLVFDITKLLLAIF 900

Qy        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
              || ::|: :| :|||||   |||:| :|| :||| |||||:: ||  ||||:|:||||:
Db        901 GPLWILQASLLKVPYFVRVQGLLRICALARKIAGGHYVQMAIIKLGALTGTYVYNHLTPL 960
```



RESULT 13
ABG32459
ID ABG32459 standard; protein; 3010 AA.
XX

AC ABG32459;
XX
DT 15-NOV-2002 (first entry)
XX
DE Hepatitis C virus Con 1 isolate polyprotein mutant #8.
XX
KW HCV; Con 1; adaptive mutation; liver failure; cirrhosis; mutant; mutein;
KW hepatocellular carcinoma; NS3; NS5; encephalomyocarditis virus; EMCV;
KW internal ribosome entry site; IRES; NS5A; HCV replication; polyprotein.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 2198
FT /note= "Wild-type Leu substituted by Ser"
XX
PN WO200259321-A2.
XX
PD 01-AUG-2002.
XX
PF 16-JAN-2002; 2002WO-EP000526.
XX
PR 23-JAN-2001; 2001US-0263479P.
XX
PA (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
XX
PI De Francesco R, Migliaccio G, Paonessa G;
XX
DR WPI; 2002-599793/64.
XX
PT New Hepatitis C virus (HCV) replicons comprising altered HCV NS3 or HCV
PT NS5 encoding region, or encephalomyocarditis virus (EMCV) internal
PT ribosome entry site (IRES) region, useful in studying HCV replication and
PT expression.
XX
PS Claim 1; Page; 69pp; English.
XX
CC The invention relates to nucleic acid molecules comprising altered HCV
CC NS3 or HCV NS5 encoding region, or encephalomyocarditis virus (EMCV)
CC internal ribosome entry site (IRES) region coding for one or more NS3,
CC NS5A, or EMCV IRES mutations, respectively. The location of the mutations
CC are detailed in the specification. Also included are (1) an expression
CC vector comprising a nucleotide sequence coding for the altered nucleic
CC acids, which is transcriptionally coupled to an exogenous promoter; (2) a
CC recombinant cell human hepatoma cell comprising the altered nucleic acids
CC ; (3) a recombinant cell produced by introducing into a human hepatoma
CC cell the altered nucleic acids; (4) producing an HCV (hepatitis C virus)
CC replicon enhanced cell or which containing a functional HCV replicon; (5)
CC an HCV replicon enhanced cells made in the method; and (6) measuring the
CC ability of a compound to affect HCV activity. The HCV replicons and HCV
CC replicon enhanced cells are useful in studying HCV replication and
CC expression, and HCV and host cell interactions, producing HCV RNA and
CC proteins, and providing a system for measuring the ability of a compound
CC to modulate one or more HCV activities e.g. to discover drugs which may
CC treat HCV mediated diseases such as liver failure, cirrhosis and
CC hepatocellular carcinoma. The present sequence is the HCV replicon Con 1
CC polyprotein (comprising the Core, E1, E2, P7, NS2, NS3, NS4A, NS4B, NS5A
CC and NS5B proteins), NS5A mutant of the invention. Note: The present
CC sequence is not shown in the specification but was created by the indexer
CC using the HCV sequence appearing as ABG32451 and the information in claim
CC 1
XX
SQ Sequence 3010 AA;

Query Match 74.7%; Score 12034.5; DB 5; Length 3010;
Best Local Similarity 71.8%; Pred. No. 0;
Matches 2183; Conservative 352; Mismatches 469; Indels 35; Gaps 7;

```
Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              ||||||| |:  |::| :|||||||||||||||||||||||||||||| ||| ||||:|
Db         61 RRQPIPKARQPEGRAWAQPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPTDPRRRSRNLG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
              |||||||||||||||||:||||||| |||||||||||||||:||||||||||||||||||
Db        121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
              ||||:| | || ||:|:|  | |||||:| || ::    ::| |||||| :  |:|:||:
Db        181 LLSCLTIPASAYEVRNVSGVYHVTNDCSNASIVYEAADMIMHTPGCVPCVRENNSSRCWV 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
               ::| :| :      | :| |:|::| :| ||||:||||||| | | ||:|  ||: |
Db        241 ALTPTLAARNASVPTTTIRRHVDLLVGAAALCSAMYVGDLCGSVFLVAQLFTFSPRRHET 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
              |||||||||| :|||||||||||||||| ::::  :|:|: ::|:::||||||: |||
Db        301 VQDCNCSIYPGHVTGHRMAWDMMMNWSPTAALVVSQLLRIPQAVVDMVAGAHWGVLAGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
              |:|| | ||||::::|| |||| |:  ||: |: | :|||| |  ||||||||||||
Db        361 YYSMVGNWAKVLIVMLLFAGVDGGTYVTGGTMAKNTLGITSLFSPGSSQKIQLVNTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
              |||||||||||||:|||:|:||| | ||||||||||::|  |:||  ||| : |  | ::
Db        421 HINRTALNCNDSLNTGFLAALFYVHKFNSSGCPERMASCSPIDAFAQGWGPITY--NESH 478

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
              | ||||||| || ||:| |  |||||||||||||||||| | |||:|||||||| |
Db        479 SSDQRPYCWHYAPRPCGIVPAAQVCGPVYCFTPSPVVVGTTDRFGVPTYSWGENETDVLL 538

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
              ||:|||| |:|||||||:|:|:||||| |||      | :  | ||||||||||: || |
Db        539 LNNTRPPQGNWFGCTWMNSTGFTKTCGGPPCNIGGIGNKT--LTCPTDCFRKHPEATYTK 596

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
              |||||||||||: ||||||||||||:||||:|||||||||| ||||:|||:||||||
Db        597 CGSGPWLTPRCLVHYPYRLWHYPCTVNFTIFKVRMYVGGVEHRLEAACNWTRGERCNLED 656

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
              ||||:||||| ||||| :||||:: |||||||:|||||:||||::||:  |:  : ::|
Db        657 RDRSELSPLLLSTTEWQVLPCSFTTLPALSTGLIHLHQNVVDVQYLYGIGSAVVSFAIKW 716

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
              |:|:|||||||||||||||::|: ||||||| ||:|:||| |  :| | |::|| |||
Db        717 EYVLLLFLLLADARVCACLWMMLLIAQAEAALENLVVLNAASVAGAHGILSFLVFFCAAW 776

Qy        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
              |||||:|| | |:| |:|   ||||||| :||| |  :    | |: | : | ||:| ||
Db        777 YIKGRLVPGAAYALYGVWPLLLLLLALPPRAYAMDREMAASCGGAVFVGLILLTLSPHYK 836

Qy        841 TLLSRPLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
              |:| :||| | :|  || :| | ||: |||||| :|      :| ::| ||| |||:|
Db        837 LFLARLIWWLQYFITRAEAHLQVWIPPLNVRGGRDAVILLTCAIHPELIFTITKILLAIL 896

Qy        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
              ||  :|:  :|:|||||||| |:| | : | :||| ||||||: |   ||||:||||||:
Db        897 GPLMVLQAGITKVPYFVRAHGLIRACMLVRKVAGGHYVQMALMKLAALTGTYVYDHLTPL 956

Qy        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
              ||| :|||||||||:|| || ||| ||||:|||||||: ||||||| ||| ||||||
Db        957 RDWAHAGLRDLAVAVEPVVFSDMETKVITWGADTAACGDIILGLPVSARRGREIHLGPAD 1016

Qy       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080
                :|| ||||||||:|||||||| |: |:|||: :  ||:||:|| ||||| | ::|
Db       1017 SLEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDRNQVEGEVQVVSTATQSFLATCVNG 1076

Qy       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
              | ||||||||:||||| :||:||||:: : ||||| :||| :|| ||||||: |||||||:
Db       1077 VCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRH 1136

Qy       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200
```

```
Db       1137 ADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFV 1196

Qy       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260

Db       1197 PVESMETTMRSPVFTDNSSPPAVPQTFQVAHLHAPTGSGKSTKVPAAYAAQGYKVLVLNP 1256

Qy       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320

Db       1257 SVAATLGFGAYMSKAHGIDPNIRTGVRTITTGAPITYSTYGKFLADGGCSGGAYDIIICD 1316

Qy       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380

Db       1317 ECHSTDSTTILGIGTVLDQAETAGARLVVLATATPPGSVTVPHPNIEEVALSSTGEIPFY 1376

Qy       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440

Db       1377 GKAIPIETIKGGRHLIFCHSKKKCDELAAKLSGLGLNAVAYYRGLDVSVIPTSGDVIVVA 1436

Qy       1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTIITTQIVPQDAVSRSQRRGRTGRGRL 1500

Db       1437 TDALMTGFTGDFDSVIDCNTCVTQTVDFSLDPTFTIETTTVPQDAVSRSQRRGRTGRGRM 1496

Qy       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560

Db       1497 GIYRFVTPGERPSGMFDSSVLCECYDAGCAWYELTPAETSVRLRAYLNTPGLPVCQDHLE 1556

Qy       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620

Db       1557 FWESVFTGLTHIDAHFLSQTKQAGDNFPYLVAYQATVCARAQAPPPSWDQMWKCLIRLKP 1616

Qy       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680

Db       1617 TLHGPTPLLYRLGAVQNEVTTTHPITKYIMACMSADLEVVTSTWVLVGGVLAALAAYCLT 1676

Qy       1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740

Db       1677 TGSVVIVGRIILSGKPAIIPDREVLYREFDEMEECASHLPYIEQGMQLAEQFKQKAIGLL 1736

Qy       1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800

Db       1737 QTATKQAEAAAPVVESKWRTLEAFWAKHMWNFISGIQYLAGLSTLPGNPAIASLMAFTAS 1796

Qy       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860

Db       1797 ITSPLTTQHTLLFNILGGWVAAQLAPPSAASAFVGAGIAGAAVGSIGLGKVLVDILAGYG 1856

Qy       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920

Db       1857 AGVAGALVAFKVMSGEMPSTEDLVNLLPAILSPGALVVGVVCAAILRRHVGPGEGAVQWM 1916

Qy       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980

Db       1917 NRLIAFASRGNHVSPTHYVPESDAAARVTQILSSLTITQLLKRLHQWINEDCSTPCSGSW 1976

Qy       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040

Db       1977 LRDVWDWICTVLTDFKTWLQSKLLPRLPGVPFFSCQRGYKGVWRGDGIMQTTCPCGAQIT 2036

Qy       2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100

Db       2037 GHVKNGSMRIVGPRTCSNTWHGTFPINAYTTGPCTPSPAPNYSRALWRVAAEEYVEVTRV 2096

Qy       2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160

Db       2097 GDFHYVTGMTTDNVKCPCQVPAPEFFTEVDGVRLHRYAPACKPLLREEVTFLVGLNQYLV 2156

Qy       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220

Db       2157 GSQLPCEPEPDVAVLTSMLTDPSHITAETAKRRLARGSPPSSASSSASQLSAPSLKATCT 2216

Qy       2221 THGKAYDVDMVDANLF----MGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYML 2276

Db       2217 TRHDSPDADLIEANLLWRQEMGGNITRVESENKVVILDSFEPLQAEEDEREVSVPAEILR 2276

Qy       2277 PKKRFPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLS 2336

Db       2277 RSRKFPRAMPIWARPDYNPPLLESWKDPDYVPPVVHGCPLPPAKAPPIPPPRRKRTVVLS 2336

Qy       2337 EDSIGDALQQLAIKSFGQPPPSG-DSGLSTGAGAADSGSQTPPDELALSETGSISSMPPL 2395

Db       2337 ESTVSSALAELATKTFGSSESSAVDSGTATA-----SPDQPSDDGDAGSDVESYSSMPPL 2391

Qy       2396 EGELGDPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEE-DDSVVCCSMSYSWTGALIT 2454

Db       2392 EGEPGDPDL----------SDGSWSTVSEEASEDVVCCSMSYTWTGALIT 2431

Qy       2455 PCSPEEEKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIK 2514

Db       2432 PCAAEETKLPINALSNSLLRHHNLVYATTSRSASLRQKKVTFDRLQVLDDHYRDVLKEMK 2491

Qy       2515 LAASKVTARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSET 2574

Db       2492 AKASTVKAKLLSVEEACKLTPPHSARSKFGYGAKDVRNLSSKAVNHIRSVWKDLLEDTET 2551

Qy       2575 PIPTTIMAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYG 2634

Db       2552 PIDTTIMAKNEVFCVQPEKGGRKPARLIVFPDLGVRVCEKMALYDVVSTLPQAVMGSSYG 2611

Qy       2635 FQYSPAQRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTA 2694

Db       2612 FQYSPGQRVEFLVNAWKAKKCPMGFAYDTRCFDSTVTENDIRVEESIYQCCDLAPEARQA 2671

Qy       2695 IHSLTERLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPT 2754

Db       2672 IRSLTERLYIGGPLTNSKGQNCGYRRCRASGVLTTSCGNTLTCYLKAAAACRAAKLQDCT 2731

Qy       2755 MLVCGDDLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSV 2814

Db       2732 MLVCGDDLVVICESAGTQEDEASLRAFTEAMTRYSAPPGDPPKPEYDLELITSCSSNVSV 2791

Qy       2815 ALGPQGRRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSI 2874

Db       2792 AHDASGKRVYYLTRDPTTPLARAAWETARHTPVNSWLGNIIMYAPTLWARMILMTHFFSI 2851

Qy       2875 LMAQDTLDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKL 2934

Db       2852 LLAQEQLEKALDCQIYGACYSIEPLDLPQIIQRLHGLSAFSLHSYSPGEINRVASCLRKL 2911

Qy       2935 GAPPLRAWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFT 2994

Db       2912 GVPPLRVWRHRARSVRARLLSQGGRAATCGKYLFNWAVRTKLKLTPIPAASQLDLSSWFV 2971

Qy       2995 VGAGGGDIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033

Db       2972 AGYSGGDIYHSLSRARPRWFMWCLLLLSVGVGIYLLPNR 3010
```

RESULT 14
ABG32455
ID ABG32455 standard; protein; 3010 AA.
XX
AC ABG32455;
XX
DT 15-NOV-2002 (first entry)
XX
DE Hepatitis C virus Con 1 isolate polyprotein mutant #4.
XX
KW HCV; Con 1; adaptive mutation; liver failure; cirrhosis; mutant; mutein;
KW hepatocellular carcinoma; NS3; NS5; encephalomyocarditis virus; EMCV;
KW internal ribosome entry site; IRES; NS5A; HCV replication; polyprotein.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
FH Key Location/Qualifiers

```
FT   Misc-difference 2041
FT                   /note= "Wild-type Asn substituted by Thr"
XX
PN   WO200259321-A2.
XX
PD   01-AUG-2002.
XX
PF   16-JAN-2002; 2002WO-EP000526.
XX
PR   23-JAN-2001; 2001US-0263479P.
XX
PA   (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
XX
PI   De Francesco R, Migliaccio G, Paonessa G;
XX
DR   WPI; 2002-599793/64.
XX
PT   New Hepatitis C virus (HCV) replicons comprising altered HCV NS3 or HCV
PT   NS5 encoding region, or encephalomyocarditis virus (EMCV) internal
PT   ribosome entry site (IRES) region, useful in studying HCV replication and
PT   expression.
XX
PS   Claim 1; Page; 69pp; English.
XX
CC   The invention relates to nucleic acid molecules comprising altered HCV
CC   NS3 or HCV NS5 encoding region, or encephalomyocarditis virus (EMCV)
CC   internal ribosome entry site (IRES) region coding for one or more NS3,
CC   NS5A, or EMCV IRES mutations, respectively. The location of the mutations
CC   are detailed in the specification. Also included are (1) an expression
CC   vector comprising a nucleotide sequence coding for the altered nucleic
CC   acids, which is transcriptionally coupled to an exogenous promoter; (2) a
CC   recombinant cell human hepatoma cell comprising the altered nucleic acids
CC   ; (3) a recombinant cell produced by introducing into a human hepatoma
CC   cell the altered nucleic acids; (4) producing an HCV (hepatitis C virus)
CC   replicon enhanced cell or which containing a functional HCV replicon; (5)
CC   an HCV replicon enhanced cells made in the method; and (6) measuring the
CC   ability of a compound to affect HCV activity. The HCV replicons and HCV
CC   replicon enhanced cells are useful in studying HCV replication and
CC   expression, and HCV and host cell interactions, producing HCV RNA and
CC   proteins, and providing a system for measuring the ability of a compound
CC   to modulate one or more HCV activities e.g. to discover drugs which may
CC   treat HCV mediated diseases such as liver failure, cirrhosis and
CC   hepatocellular carcinoma. The present sequence is the HCV replicon Con 1
CC   polyprotein (comprising the Core, E1, E2, P7, NS2, NS3, NS4A, NS4B, NS5A
CC   and NS5B proteins), NS5A mutant of the invention. Note: The present
CC   sequence is not shown in the specification but was created by the indexer
CC   using the HCV sequence appearing as ABG32451 and the information in claim
CC   1
XX
SQ   Sequence 3010 AA;

  Query Match             74.6%;  Score 12033.5;  DB 5;  Length 3010;
  Best Local Similarity   71.8%;  Pred. No. 0;
  Matches 2183;  Conservative  352;  Mismatches  469;  Indels   35;  Gaps    7;

Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              |||||||  |:  |::| :|||||||||||||||||||||||||||||| ||| ||||:|
Db         61 RRQPIPKARQPEGRAWAQPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPTDPRRRSRNLG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
              ||||||||||||||||:|||||| ||||||||||||||:|||||||||||||||||||||
Db        121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
              ||||:| | || ||:|:|  | |||||:| || ::    ::| |||||| :  |:|:||:
Db        181 LLSCLTIPASAYEVRNVSGVYHVTNDCSNASIVYEAADMIMHTPGCVPCVRENNSSRCWV 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
               ::| :| :     |  :| |:|::| :| ||||:||||||| | | ||:|  ||: |

Db        241 ALTPTLAARNASVPTTTIRRHVDLLVGAAALCSAMYVGDLCGSVFLVAQLFTFSPRRHET 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
              |||||||||||  :|||||||||||||||||| ::::  :|:|: ::|:::||||||: |||
Db        301 VQDCNCSIYPGHVTGHRMAWDMMMNWSPTAALVVSQLLRIPQAVVDMVAGAHWGVLAGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
              |:|| | ||||::::|| |||| |:  ||: |: | :|||| | ||||||||||||
Db        361 YYSMVGNWAKVLIVMLLFAGVDGGTYVTGGTMAKNTLGITSLFSPGSSQKIQLVNTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
              ||||||||||||:|||:|:||| | ||||||||::|  |:||  ||| : |  | ::
Db        421 HINRTALNCNDSLNTGFLAALFYVHKFNSSGCPERMASCSPIDAFAQGWGPITY--NESH 478

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
              | ||||||| || ||:| | |||||||||||||||||||| | |||:||||||| |
Db        479 SSDQRPYCWHYAPRPCGIVPAAQVCGPVYCFTPSPVVVGTTDRFGVPTYSWGENETDVLL 538

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
              ||:|||| |:|||||||||:|:||||| |||       | :  | ||||||||||: || |
Db        539 LNNTRPPQGNWFGCTWMNSTGFTKTCGGPPCNIGGIGNKT--LTCPTDCFRKHPEATYTK 596

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
              |||||||||||: |||||||||||:||||:|||||||||| ||||:|||:||||||
Db        597 CGSGPWLTPRCLVHYPYRLWHYPCTVNFTIFKVRMYVGGVEHRLEAACNWTRGERCNLED 656

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
              ||||:|||||  |||||  :||||:: |||||||:|||||:||||::||:  |:  : ::|
Db        657 RDRSELSPLLLSTTEWQVLPCSFTTLPALSTGLIHLHQNVVDVQYLYGIGSAVVSFAIKW 716

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
              |:|:|||||||||||||||::|: |||||| ||:|:||| |  :| | |::|| |||
Db        717 EYVLLLFLLLADARVCACLWMMLLIAQAEAALENLVVLNAASVAGAHGILSFLVFFCAAW 776

Qy        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
              |||||:|| | |:| |:|    ||||||| :||| | :    | |: | : | ||:| ||
Db        777 YIKGRLVPGAAYALYGVWPLLLLLLALPPRAYAMDREMAASCGGAVFVGLILLTLSPHYK 836

Qy        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
              |:| :||| | :|  || :| | ||: |||||| :|      :| ::| ||| |||:|
Db        837 LFLARLIWWLQYFITRAEAHLQVWIPPLNVRGGRDAVILLTCAIHPELIFTITKILLAIL 896

Qy        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
              ||  :|:  :|:|||||||| |:| | : | :||| ||||||: |   ||||:|||||:
Db        897 GPLMVLQAGITKVPYFVRAHGLIRACMLVRKVAGGHYVQMALMKLAALTGTYVYDHLTPL 956

Qy        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
              ||| :|||||||||::|| || ||| |||:|||||||: ||||||| |||: |||||
Db        957 RDWAHAGLRDLAVAVEPVVFSDMETKVITWGADTAACGDIILGLPVSARRGREIHLGPAD 1016

Qy       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080
                :|| |||||||:|||||||| |: |:|||: :  ||:||:|| ||||| | ::|
Db       1017 SLEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDRNQVEGEVQVVSTATQSFLATCVNG 1076

Qy       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
              | ||:|||||:|||| :|||:|||:: : ||||| :||| :|| |||||: |||||||:
Db       1077 VCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRH 1136

Qy       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200
              ||||| ||||| ||:||||||:| |||||||:||| |||||:||||||:|||||::||:
Db       1137 ADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFV 1196

Qy       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260
              |||:::   ||| |:|||:|||||||:|| :|||||||||||| ||||||||||||
Db       1197 PVESMETTMRSPVFTDNSSPPAVPQTFQVAHLHAPTGSGKSTKVPAAYAAQGYKVLVLNP 1256

Qy       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320
              |||||||||:|||||:|||||||:|||||||||||||||||||:|||||||||||||||
Db       1257 SVAATLGFGAYMSKAHGIDPNIRTGVRTITTGAPITYSTYGKFLADGGCSGGAYDIIICD 1316

Qy       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380
              |||: |||||||||||||||| || |||||||||| ||||||||||  ||||||
Db       1317 ECHSTDSTTILGIGTVLDQAETAGARLVVLATATPPGSVTVPHPNIEEVALSSTGEIPFY 1376
```

```
Qy       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440
Db       1377 GKAIPIETIKGGRHLIFCHSKKKCDELAAKLSGLGLNAVAYYRGLDVSVIPTSGDVIVVA 1436

Qy       1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTIITTQIVPQDAVSRSQRRGRTGRGRL 1500
Db       1437 TDALMTGFTGDFDSVIDCNTCVTQTVDFSLDPTFTIETTTVPQDAVSRSQRRGRTGRGRM 1496

Qy       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560
Db       1497 GIYRFVTPGERPSGMFDSSVLCECYDAGCAWYELTPAETSVRLRAYLNTPGLPVCQDHLE 1556

Qy       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
Db       1557 FWESVFTGLTHIDAHFLSQTKQAGDNFPYLVAYQATVCARAQAPPPSWDQMWKCLIRLKP 1616

Qy       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680
Db       1617 TLHGPTPLLYRLGAVQNEVTTTHPITKYIMACMSADLEVVTSTWVLVGGVLAALAAYCLT 1676

Qy       1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740
Db       1677 TGSVVIVGRIILSGKPAIIPDREVLYREFDEMEECASHLPYIEQGMQLAEQFKQKAIGLL 1736

Qy       1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
Db       1737 QTATKQAEAAAPVVESKWRTLEAFWAKHMWNFISGIQYLAGLSTLPGNPAIASLMAFTAS 1796

Qy       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860
Db       1797 ITSPLTTQHTLLFNILGGWVAAQLAPPSAASAFVGAGIAGAAVGSIGLGKVLVDILAGYG 1856

Qy       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920
Db       1857 AGVAGALVAFKVMSGEMPSTEDLVNLLPAILSPGALVVGVVCAAILRRHVGPGEGAVQWM 1916

Qy       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980
Db       1917 NRLIAFASRGNHVSPTHYVPESDAAARVTQILSSLTITQLLKRLHQWINEDCSTPCSGSW 1976

Qy       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040
Db       1977 LRDVWDWICTVLTDFKTWLQSKLLPRLPGVPFFSCQRGYKGVWRGDGIMQTTCPCGAQIT 2036

Qy       2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100
Db       2037 GHVKTGSMRIVGPRTCSNTWHGTFPINAYTTGPCTPSPAPNYSRALWRVAAEEYVEVTRV 2096

Qy       2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160
Db       2097 GDFHYVTGMTTDNVKCPCQVPAPEFFTEVDGVRLHRYAPACKPLLREEVTFLVGLNQYLV 2156

Qy       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220
Db       2157 GSQLPCEPEPDVAVLTSMLTDPSHITAETAKRRLARGSPPSLASSSASQLSAPSLKATCT 2216

Qy       2221 THGKAYDVDMVDANLF----MGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYML 2276
Db       2217 TRHDSPDADLIEANLLWRQEMGGNITRVESENKVVILDSFEPLQAEEDEREVSVPAEILR 2276

Qy       2277 PKKRFPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLS 2336
Db       2277 RSRKFPRAMPIWARPDYNPPLLESWKDPDYVPPVVHGCPLPPAKAPPIPPPRRKRTVVLS 2336

Qy       2337 EDSIGDALQQLAIKSFGQPPPSG-DSGLSTGAGAADSGSQTPPDELALSETGSISSMPPL 2395
Db       2337 ESTVSSALAELATKTFGSSESSAVDSGTATA-----SPDQPSDDGDAGSDVESYSSMPPL 2391

Qy       2396 EGELGDPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEE-DDSVVCCSMSYSWTGALIT 2454
Db       2392 EGEPGDPDL-------------------SDGSWSTVSEEASEDVVCCSMSYTWTGALIT 2431

Qy       2455 PCSPEEEKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIK 2514
Db       2432 PCAAEETKLPINALSNSLLRHHNLVYATTSRSASLRQKKVTFDRLQVLDDHYRDVLKEMK 2491

Qy       2515 LAASKVTARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSET 2574
Db       2492 AKASTVKAKLLSVEEACKLTPPHSARSKFGYGAKDVRNLSSKAVNHIRSVWKDLLEDTET 2551

Qy       2575 PIPTTIMAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYG 2634
Db       2552 PIDTTIMAKNEVFCVQPEKGGRKPARLIVFPDLGVRVCEKMALYDVVSTLPQAVMGSSYG 2611

Qy       2635 FQYSPAQRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTA 2694
Db       2612 FQYSPGQRVEFLVNAWKAKKCPMGFAYDTRCFDSTVTENDIRVEESIYQCCDLAPEARQA 2671

Qy       2695 IHSLTERLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPT 2754
Db       2672 IRSLTERLYIGGPLTNSKGQNCGYRRCRASGVLTTSCGNTLTCYLKAAAACRAAKLQDCT 2731

Qy       2755 MLVCGDDLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSV 2814
Db       2732 MLVCGDDLVVICESAGTQEDEASLRAFTEAMTRYSAPPGDPPKPEYDLELITSCSSNVSV 2791

Qy       2815 ALGPQGRRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSI 2874
Db       2792 AHDASGKRVYYLTRDPTTPLARAAWETARHTPVNSWLGNIIMYAPTLWARMILMTHFFSI 2851

Qy       2875 LMAQDTLDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKL 2934
Db       2852 LLAQEQLEKALDCQIYGACYSIEPLDLPQIIQRLHGLSAFSLHSYSPGEINRVASCLRKL 2911

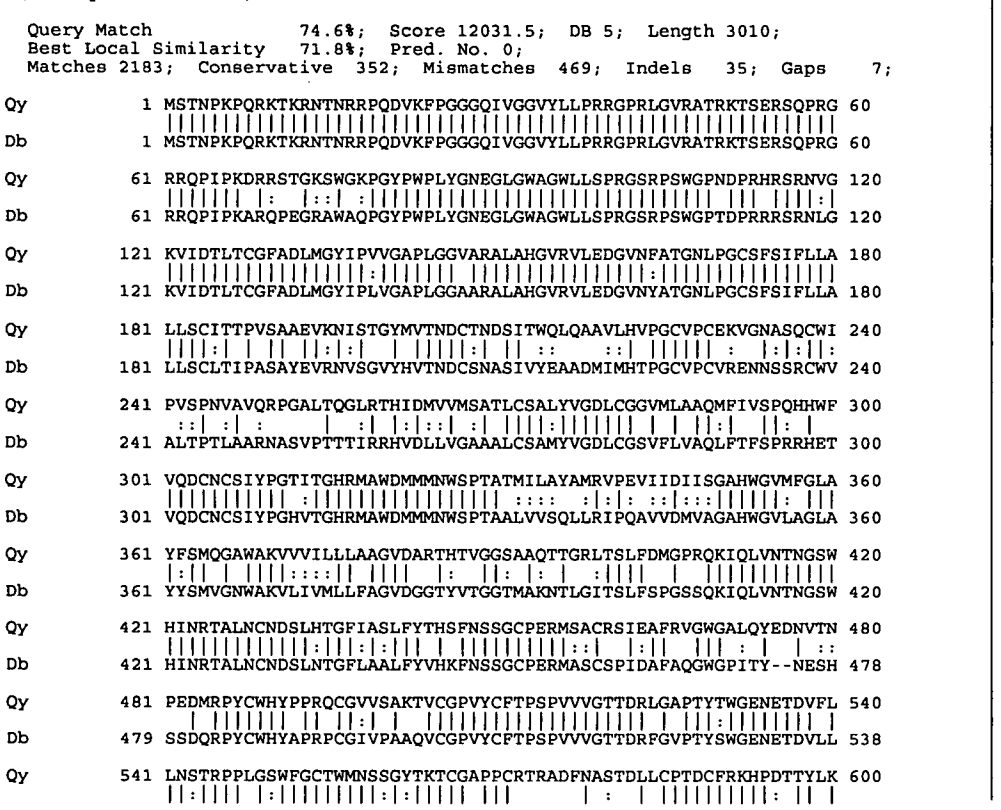
Qy       2935 GAPPLRAWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFT 2994
Db       2912 GVPPLRVWRHRARSVRARLLSQGGRAATCGKYLFNWAVRTKLKLTPIPAASQLDLSSWFV 2971

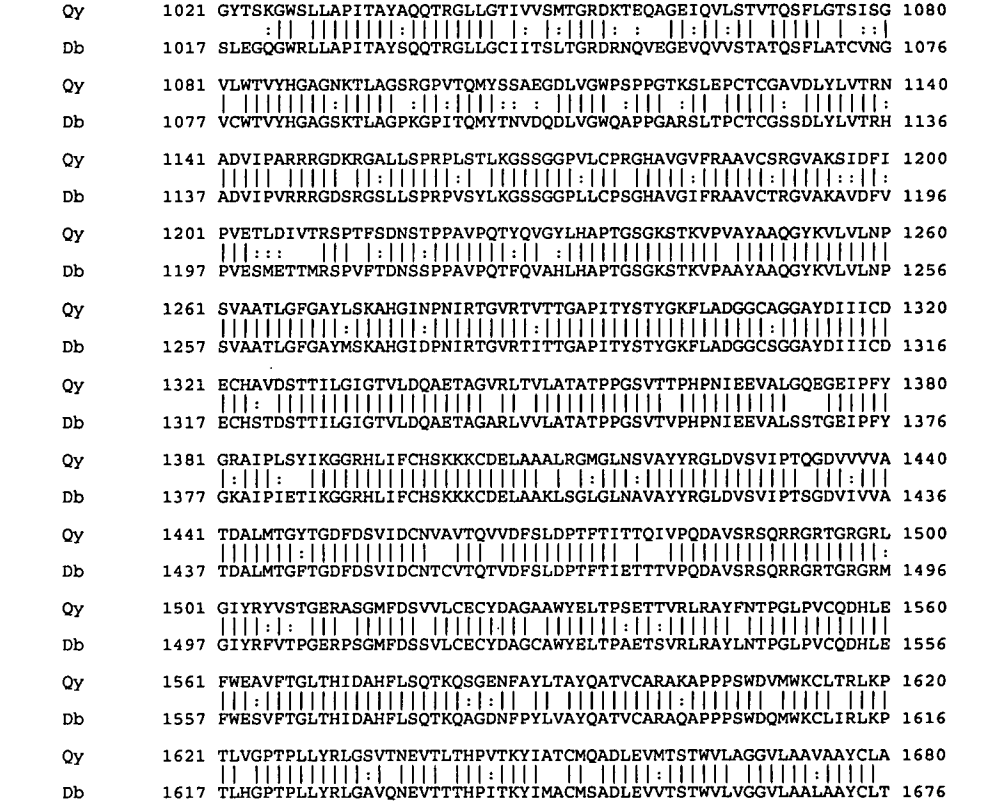
Qy       2995 VGAGGGDIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
Db       2972 AGYSGGDIYHSLSRARPRWFMWCLLLLSVGVGIYLLPNR 3010
```

RESULT 15
ABG32451
ID ABG32451 standard; protein; 3010 AA.
XX
AC ABG32451;
XX
DT 15-NOV-2002 (first entry)
XX
DE Hepatitis C virus Con 1 isolate polyprotein.
XX
KW HCV; Con 1; adaptive mutation; liver failure; cirrhosis;
KW hepatocellular carcinoma; NS3; NS5; encephalomyocarditis virus; EMCV;
KW internal ribosome entry site; IRES; NS5A; HCV replication; polyprotein.
XX
OS Hepatitis C virus.
XX
PN WO200259321-A2.
XX
PD 01-AUG-2002.
XX
PF 16-JAN-2002; 2002WO-EP000526.
XX
PR 23-JAN-2001; 2001US-0263479P.
XX
PA (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
XX
PI De Francesco R, Migliaccio G, Paonessa G;
XX
DR WPI; 2002-599793/64.
DR N-PSDB; ABK91411.
XX
PT New Hepatitis C virus (HCV) replicons comprising altered HCV NS3 or HCV

PT NS5 encoding region, or encephalomyocarditis virus (EMCV) internal
PT ribosome entry site (IRES) region, useful in studying HCV replication and
PT expression.
XX
PS Claim 1; Page 34-36; 69pp; English.
XX
CC The invention relates to nucleic acid molecules comprising altered HCV
CC NS3 or HCV NS5 encoding region, or encephalomyocarditis virus (EMCV)
CC internal ribosome entry site (IRES) region coding for one or more NS3,
CC NS5A, or EMCV IRES mutations, respectively. The location of the mutations
CC are detailed in the specification. Also included are (1) an expression
CC vector comprising a nucleotide sequence coding for the altered nucleic
CC acids, which is transcriptionally coupled to an exogenous promoter; (2) a
CC recombinant cell human hepatoma cell comprising the altered nucleic acids
CC ; (3) a recombinant cell produced by introducing into a human hepatoma
CC cell the altered nucleic acids; (4) producing an HCV (hepatitis C virus)
CC replicon enhanced cell or which containing a functional HCV replicon; (5)
CC an HCV replicon enhanced cells made in the method; and (6) measuring the
CC ability of a compound to affect HCV activity. The HCV replicons and HCV
CC replicon enhanced cells are useful in studying HCV replication and
CC expression, and HCV and host cell interactions, producing HCV RNA and
CC proteins, and providing a system for measuring the ability of a compound
CC to modulate one or more HCV activities e.g. to discover drugs which may
CC treat HCV mediated diseases such as liver failure, cirrhosis and
CC hepatocellular carcinoma. The present sequence is the HCV replicon Con 1
CC polyprotein (comprising the Core, E1, E2, P7, NS2, NS3, NS4A, NS4B, NS5A
CC and NS5B proteins) used as a basis for the adaptive mutations of the
CC invention
XX
SQ Sequence 3010 AA;

Query Match 74.6%; Score 12031.5; DB 5; Length 3010;
Best Local Similarity 71.8%; Pred. No. 0;
Matches 2183; Conservative 352; Mismatches 469; Indels 35; Gaps 7;





```
Qy       1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740
Db       1677 TGSVVIVGRIILSGKPAIIPDREVLYREFDEMEECASHLPYIEQGMQLAEQFKQKAIGLL 1736

Qy       1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
Db       1737 QTATKQAEAAAPVVESKWRTLEAFWAKHMWNFISGIQYLAGLSTLPGNPAIASLMAFTAS 1796

Qy       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860
Db       1797 ITSPLTTQHTLLFNILGGWVAAQLAPPSAASAFVGAGIAGAAVGSIGLGKVLVDILAGYG 1856

Qy       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920
Db       1857 AGVAGALVAFKVMSGEMPSTEDLVNLLPAILSPGALVVGVVCAAILRRHVGPGEGAVQWM 1916

Qy       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980
Db       1917 NRLIAFASRGNHVSPTHYVPESDAAARVTQILSSLTITQLLKRLHQWINEDCSTPCSGSW 1976

Qy       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040
Db       1977 LRDVWDWICTVLTDFKTWLQSKLLPRLPGVPFFSCQRGYKGVWRGDGIMQTTCPCGAQIT 2036

Qy       2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100
Db       2037 GHVKNGSMRIVGPRTCSNTWHGTFPINAYTTGPCTPSPAPNYSRALWRVAAEEYVEVTRV 2096

Qy       2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160
Db       2097 GDFHYVTGMTTDNVKCPCQVPAPEFFTEVDGVRLHRYAPACKPLLREEVTFLVGLNQYLV 2156

Qy       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220
Db       2157 GSQLPCEPEPDVAVLTSMLTDPSHITAETAKRRLARGSPPSLASSSASQLSAPSLKATCT 2216

Qy       2221 THGKAYDVDMVDANLF----MGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYML 2276
Db       2217 TRHDSPDADLIEANLLWRQEMGGNITRVESENKVVILDSFEPLQAEEDEREVSVPAEILR 2276

Qy       2277 PKKRFPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLS 2336
Db       2277 RSRKFPRAMPIWARPDYNPPLLESWKDPDYVPPVVHGCPLPPAKAPPIPPPRRKRTVVLS 2336

Qy       2337 EDSIGDALQQLAIKSFGQPPPSG-DSGLSTGAGAADSGSQTPPDELALSETGSISSMPPL 2395
Db       2337 ESTVSSALAELATKTFGSSESSAVDSGTATA-----SPDQPSDDGDAGSDVESYSSMPPL 2391

Qy       2396 EGELGDPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEE-DDSVVCCSMSYSWTGALIT 2454
Db       2392 EGEPGDPDL--------------------SDGSWSTVSEEASEDVVCCSMSYTWTGALIT 2431

Qy       2455 PCSPEEEKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIK 2514
Db       2432 PCAAEETKLPINALSNSLLRHHNLVYATTSRSASLRQKKVTFDRLQVLDDHYRDVLKEMK 2491

Qy       2515 LAASKVTARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSET 2574
Db       2492 AKASTVKAKLLSVEEACKLTPPHSARSKFGYGAKDVRNLSSKAVNHIRSVWKDLLEDTET 2551

Qy       2575 PIPTTIMAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYG 2634
Db       2552 PIDTTIMAKNEVFCVQPEKGGRKPARLIVFPDLGVRVCEKMALYDVVSTLPQAVMGSSYG 2611

Qy       2635 FQYSPAQRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTA 2694
Db       2612 FQYSPGQRVEFLVNAWKAKKCPMGFAYDTRCFDSTVTENDIRVEESIYQCCDLAPEARQA 2671

Qy       2695 IHSLTERLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPT 2754
Db       2672 IRSLTERLYIGGPLTNSKGQNCGYRRCRASGVLTTSCGNTLTCYLKAAAACRAAKLQDCT 2731

Qy       2755 MLVCGDDLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSV 2814
Db       2732 MLVCGDDLVVICESAGTQEDEASLRAFTEAMTRYSAPPGDPPKPEYDLELITSCSSNVSV 2791

Qy       2815 ALGPQGRRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSI 2874
Db       2792 AHDASGKRVYYLTRDPTTPLARAAWETARHTPVNSWLGNIIMYAPTLWARMILMTHFFSI 2851

Qy       2875 LMAQDTLDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKL 2934
Db       2852 LLAQEQLEKALDCQIYGACYSIEPLDLPQIIQRLHGLSAFSLHSYSPGEINRVASCLRKL 2911

Qy       2935 GAPPLRAWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFT 2994
Db       2912 GVPPLRVWRHRARSVRARLLSQGGRAATCGKYLFNWAVRTKLKLTPIPAASQLDLSSWFV 2971

Qy       2995 VGAGGGDIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
Db       2972 AGYSGGDIYHSLSRARPRWFMWCLLLLSVGVGIYLLPNR 3010
```

Search completed: October 28, 2005, 15:59:04
Job time : 304 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2005, 15:48:07 ; Search time 75 Seconds
(without alignments)
3891.006 Million cell updates/sec

Title: US-09-980-559-2
Perfect score: 16120
Sequence: 1 MSTNPKPQRKTKRNTNRRPQ.....LLFGLLLLFVGVGLFLLPAR 3033

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	15830	98.2	3033	1	JQ1303	genome polyprotein
2	13971	86.7	3033	1	GNWVJ8	genome polyprotein
3	11998.5	74.4	3010	1	GNWVCJ	genome polyprotein
4	11984.5	74.3	3010	1	GNWVTC	genome polyprotein
5	11979.5	74.3	3010	1	A45573	genome polyprotein
6	11922	74.0	3011	1	GNWVC3	genome polyprotein
7	11911.5	73.9	3010	1	GNWVTW	genome polyprotein
8	11895	73.8	3011	1	S40770	genome polyprotein
9	11843.5	73.5	3010	1	S18030	genome polyprotein
10	11796	73.2	3011	1	GNWVCH	genome polyprotein
11	11500.5	71.3	3014	1	JC5620	genome polyprotein
12	3490.5	21.7	874	2	JQ0881	genome polyprotein
13	3263	20.2	876	2	PC2219	polypeptide - hepa
14	3231	20.0	782	2	S19875	genome polyprotein
15	3214	19.9	782	2	S19876	genome polyprotein
16	3194	19.8	782	2	S18032	genome polyprotein
17	3194	19.8	782	2	S18031	genome polyprotein
18	3126	19.4	787	2	PN0677	hypothetical prote
19	2937.5	18.2	874	2	JQ0883	genome polyprotein
20	2671	16.6	640	2	JQ1584	genome polyprotein
21	2596	16.1	716	2	JQ1366	polyprotein - hepa
22	2199	13.6	3005	2	T08841	polyprotein - dour
23	2114.5	13.1	2970	2	T08839	polyprotein - marm
24	2111.5	13.1	520	2	JQ1925	polyprotein - hepa
25	2108.5	13.1	523	2	JQ1926	polyprotein - hepa
26	2099	13.0	513	2	PC1284	genome polyprotein
27	2089	13.0	550	2	JH0711	genome polyprotein
28	2074	12.9	492	2	PS0326	polyprotein - hepa
29	2056	12.8	513	2	A44150	structural protein
30	1958.5	12.1	492	2	S41288	genome polyprotein
31	1808	11.2	441	2	S12707	genome polyprotein
32	1761	10.9	350	2	S35631	genome polyprotein
33	1734.5	10.8	386	2	S68016	ATPase/RNA helicas
34	1630	10.1	411	2	PC2061	genome polyprotein
35	1622	10.1	411	2	PC2060	genome polyprotein
36	1486.5	9.2	369	2	S21471	genome polyprotein
37	1451	9.0	415	2	PC4407	envelope protein -
38	1450	9.0	365	2	JQ0880	NS5 protein - hepa
39	1430	8.9	365	2	JQ0879	NS5 protein - hepa
40	1376	8.5	322	2	JN0265	genome polyprotein
41	1085	6.7	315	2	PS0165	envelope glycoprot
42	1076	6.7	315	2	PN0011	envelope glycoprot
43	1062	6.6	315	2	PS0164	envelope glycoprot
44	984	6.1	216	2	S21337	genome polyprotein
45	962	6.0	234	2	S32742	genome polyprotein

ALIGNMENTS

RESULT 1
JQ1303
genome polyprotein - hepatitis C virus (isolate HC-J6)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 09-Jul-2004
C;Accession: JQ1303
R;Okamoto, H.; Okada, S.; Sugiyama, Y.; Kurai, K.; Iizuka, H.; Machida, A.; Miyakawa, Y.
J. Gen. Virol. 72, 2697-2704, 1991
A;Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a hum
A;Reference number: JQ1303; MUID:92044440; PMID:1658196
A;Accession: JQ1303
A;Molecule type: genomic RNA
A;Residues: 1-3033 <OKA>
A;Cross-references: UNIPROT:P26660; GB:D00944; NID:g221650; PIDN:BAA00792.1; PID:g221651
A;Experimental source: isolate HC-J6 from a Japanese individual
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; glycoprotein; hydrolase; P-loop; polyprotein; serine proteinase; transm
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-733/Product: nonstructural protein NS1 #status predicted <NS1>
F;734-1010/Product: nonstructural protein NS2 #status predicted <NS2>
F;1011-1619/Product: hepacivirin #status predicted <NS3>
F;1316-1321/Region: nucleotide-binding motif B
F;1320-1323/Region: DEXH motif
F;1620-1866/Product: nonstructural protein NS4a #status predicted <N4A>
F;1867-2017/Product: nonstructural protein NS4b #status predicted <N4B>
F;2018-3033/Product: nonstructural protein NS5 #status predicted <NO5>
F;196,209,234,305,325,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,28

Query Match 98.2%; Score 15830; DB 1; Length 3033;
Best Local Similarity 97.8%; Pred. No. 0;
Matches 2967; Conservative 30; Mismatches 36; Indels 0; Gaps 0;

```
Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| |:|||
Db        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNTSRCWI 240
```

```
Qy       241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
Db       241 PVSPNVAVQQPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300

Qy       301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
Db       301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIIGGAHWGVMFGLA 360

Qy       361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
Db       361 YFSMQGAWAKVVVILLLAAGVDAQTHTVGGSTAHNARTLTGMFSLGARQKIQLINTNGSW 420

Qy       421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
Db       421 HINRTALNCNDSLHTGFLASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480

Qy       481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
Db       481 PEDMRPYCWHYPPRQCGVVSASSVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540

Qy       541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
Db       541 LNSTRPPQGSWFGCTWMNSTGYTKTCGAPPCRIRADFNASMDLLCPTDCFRKHPDTTYIK 600

Qy       601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
Db       601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660

Qy       661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
Db       661 RDRSQLSPLLHSTTEWAILPCTYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720

Qy       721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
Db       721 EWVVLLFLLLADARVCACLWMLILLGQAEAALEKLVVLHAASAASCNGFLYFVIFFVAAW 780

Qy       781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
Db       781 YIKGRVVPLATYSLTGLWSFGLLLLALPQQAYAYDASVHGQIGAALLVLITLFTLTPGYK 840

Qy       841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
Db       841 TLLSRFLWWLCYLLTLAEAMVQEWAPPMQVRGGRDGIIWAVAIFCPGVVFDITKWLLAVL 900

Qy       901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
Db       901 GPAYLLKGALTRVPYFVRAHALLRMCTMVRHLAGGRYVQMVLLALGRWTGTYIYDHLTPM 960

Qy       961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
Db       961 SDWAANGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020

Qy      1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080
Db      1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTTISG 1080

Qy      1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
Db      1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140

Qy      1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200
Db      1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200

Qy      1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260
Db      1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260

Qy      1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320
Db      1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320

Qy      1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380
Db      1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380

Qy      1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440
Db      1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNAVAYYRGLDVSVIPTQGDVVVVA 1440

Qy      1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500
Db      1441 TDALMTGFTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQTVPQDAVSRSQRRGRTGRGRL 1500

Qy      1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560
Db      1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPAETTVRLRAYFNTPGLPVCQDHLE 1560

Qy      1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
Db      1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620

Qy      1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680
Db      1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680

Qy      1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740
Db      1681 TGCVCIIGRLHVNQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740

Qy      1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
Db      1741 QQASKQAQDIQPAVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800

Qy      1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860
Db      1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860

Qy      1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920
Db      1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920

Qy      1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980
Db      1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCSGSW 1980

Qy      1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040
Db      1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFISCQKGYKGVWAGTGIMTTRCPCGANIS 2040

Qy      2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100
Db      2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKIAIWRVAASEYAEVTQH 2100

Qy      2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160
Db      2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPIPKPFFRDEVSFCVGLNSFVV 2160

Qy      2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220
Db      2161 GSQLPCDPEPDTDVLTSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220

Qy      2221 THGKAYDVDMVDANLFMGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYMLPKKR 2280
Db      2221 THGKAYDVDMVDANLFMGGDVTRIESESKVVVLDSLDPMVEERSDLEPSIPSEYMLPKKR 2280

Qy      2281 FPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLSEDSI 2340
Db      2281 FPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPKKTPTPPPRRRRTVGLSESSI 2340

Qy      2341 GDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTPPDELALSETGSISSMPPLEGELG 2400
Db      2341 ADALQQLAIKSFGQPPPSGDSGLSTGADAADSGSRTPPDELALSETGSISSMPPLEGEPG 2400

Qy      2401 DPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460
```

```
Db       2401 DPDLEPEQVELQPPPQGGVVTPGSGSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460

Qy       2461 EKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIKLAASKV 2520
             ||||||||||||||||||||||:||||||||||||||| ||::|||||||||||||||
Db       2461 EKLPINPLSNSLLRYHNKVYCTTSKSASLRAKKVTFDRMQALDAHYDSVLKDIKLAASKV 2520

Qy       2521 TARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSETPIPTTI 2580
             |||||||:|||||||||||||||||||||||||||||||||||||||||||::|||||||
Db       2521 TARLLTLEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDTQTPIPTTI 2580

Qy       2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPA 2640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPA 2640

Qy       2641 QRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTE 2700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2641 QRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTE 2700

Qy       2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPTMLVCGD 2760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPTMLVCGD 2760

Qy       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820

Qy       2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDT 2880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDT 2880

Qy       2881 LDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2881 LDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2940

Qy       2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000

Qy       3001 DIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
             |||||||||||||| ||||||||||||||||
Db       3001 DIYHSVSRARPRLLLLGLLLLFVGVGLFLLPAR 3033


RESULT 2
GNWVJ8
genome polyprotein - hepatitis C virus (strain HC-J8)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 09-Jul-2004
C;Accession: A40250; PQ0397; PQ0559
R;Okamoto, H.; Kurai, K.; Okada, S.I.; Yamamoto, K.; Lizuka, H.; Tanaka, T.; Fukuda, S.;
Virology 188, 331-341, 1992
A;Title: Full-length sequence of a hepatitis C virus genome having poor homology to repc
A;Reference number: A40250; MUID:92230232; PMID:1314459
A;Accession: A40250
A;Molecule type: genomic RNA
A;Residues: 1-3033 <OKA>
A;Cross-references: UNIPROT:P26661; GB:D10988; GB:D01221; NID:g221608; PIDN:BAA01761.1;
R;Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.L
J. Gen. Virol. 73, 1131-1141, 1992
A;Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e
A;Reference number: PQ0393; MUID:92268871; PMID:1316939
A;Accession: PQ0397
A;Molecule type: genomic RNA
A;Residues: 2678-2754 <CHA>
A;Cross-references: DDBJ:D10134
A;Experimental source: isolate E-b12
R;Kato, N.; Ootsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Mori, S.; Hijikata, M.; Shimotohnc
Biochem. Biophys. Res. Commun. 181, 279-285, 1991
A;Title: Distribution of plural HCV types in Japan.
A;Reference number: PQ0554; MUID:92068204; PMID:1720309
A;Accession: PQ0559
A;Molecule type: mRNA
A;Residues: 2678-2729 <KAT>
A;Cross-references: GB:D10562; GB:D90518; NID:g221523; PIDN:BAA01418.1; PID:g221524
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
F;1-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-733/Product: nonstructural protein NS1 #status predicted <NS1>
F;734-1010/Product: nonstructural protein NS2 #status predicted <NS2>
F;1011-1619/Product: hepacivirin #status predicted <NS3>
F;1234-1241/Region: nucleotide-binding motif A (P-loop)
F;1316-1321/Region: nucleotide-binding motif B
F;1320-1323/Region: DEXH motif
F;1620-1866/Product: nonstructural protein NS4a #status predicted <N4A>
F;1867-2017/Product: nonstructural protein NS4b #status predicted <N4B>
F;2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,233,299,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,23

  Query Match            86.7%;  Score 13971;  DB 1;  Length 3033;
  Best Local Similarity  83.9%;  Pred. No. 0;
  Matches 2544;  Conservative  245;  Mismatches  244;  Indels    0;  Gaps    0;

Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              |||||||||||||||||||||||||||||||||||||||||||||:|||||||||:|
Db         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGCGWAGWLLSPRGSRPTWGPTDPRHRSRNLG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
              :||||:|||||||||||||||:||||||||||||||||:|:||||||||||||||||
Db        121 RVIDTITCGFADLMGYIPVVGAPVGGVARALAHGVRVLEDGINYATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
              ||||:| |||| ||:|||: |  ||||:|:|||||  ||||:|||||||       |||
Db        181 LLSCVTVPVSAVEVRNISSSYYATNDCSNNSITWQLTDAVLHLPGCVPCENDNGTLHCWI 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
              |:|||||:  |||| ||||:|||:||:|| |||||||||:|| ||: :| |:|||| | |
Db        241 QVTPNVAVKHRGALTRSLRTHVDMIVMAATACSALYVGDVCGAVMILSQAFMVSPQRHNF 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
              |:|||||| | |||||||||::|||| |||||| |||:::||  | ||||:||||
Db        301 TQECNCSIYQGHITGHRMAWDMMLSWSPTLTMILAYAARVPELVLEIIFGGHWGVVFGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
              ||||||||||: |||| ||||| |:: |  | :|      ||  | :| : |:||||||
Db        361 YFSMQGAWAKVIAILLLVAGVDATTYSSGQEAGRTVAGFAGLFTTGAKQNLYLINTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
              |||||||||||| |||:|||||| |||||||||:|:|| :: ||:||| |:|| ||||
Db        421 HINRTALNCNDSLQTGFLASLFYTHKFNSSGCPERLSSCRGLDDFRIGWGTLEYETNVTN 480

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
              |||||||||||| ||:| |:|||||||||||||||||: | ||||||||||||||||
Db        481 DGDMRPYCWHYPPRPCGIVPARTVCGPVYCFTPSPVVVGTTDKQGVPTYTWGENETDVFL 540

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
              ||||||| |:|||||||| :|:|||||||||| | |:|:: |||||||||||||| ||||
Db        541 LNSTRPPRGAWFGCTWMNGTGFTKTCGAPPCRIRKDYNSTIDLLCPTDCFRKHPDATYLK 600

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
              ||:|||||||:|||||||||||:|||| |||||||||| :|||||||||| |||
Db        601 CGAGPWLTPRCLVDYPYRLWHYPCTVNFTIFKARMYVGGVEHRFSAACNFTRGDRCRLED 660

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
              ||| | |||||||||||:||||:|||||||||||||||||::|||||||:|||:|
Db        661 RDRGQQSPLLHSTTEWAVLPCSFSDLPALSTGLLHLHQNIVDVQYLYGLSPALTRYIVKW 720

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
              |||||||||||||:||||||||:||||||||||:|||:||||| || |:| ||| |||
```



```
Qy       2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000
Db       2941 AWKSRARAVRASLIAQGARAAICGRYLFNWAVKTKLKLTPLPEASRLDLSGWFTVGAGGG 3000

Qy       3001 DIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
Db       3001 DIYHSVSHARPRLLLLCLLLLSVGVGIFLLPAR 3033


RESULT 3
GNWVCJ
genome polyprotein - hepatitis C virus (strain J)
N;Contains: capsid protein C; envelope protein M; major envelope protein E; nonstructura
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 09-Jul-2004
C;Accession: A39253; PS0086
R;Kato, N.; Hijikata, M.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; Shimot
Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990
A;Title: Molecular cloning of the human hepatitis C virus genome from Japanese patients
A;Reference number: A39253; MUID:91088550; PMID:2175903
A;Accession: A39253
A;Molecule type: genomic RNA
A;Residues: 1-3010 <KAT>
A;Cross-references: UNIPROT:P26662; GB:D90208; NID:g221610; PIDN:BAA14233.1; PID:g221611
R;Kato, N.; Ohkoshi, S.; Shimotohno, K.
Proc. Jpn. Acad. 65B, 219-223, 1989
A;Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence vari
A;Reference number: PS0085
A;Accession: PS0086
A;Molecule type: genomic RNA
A;Residues: 2650-2707 <KA2>
A;Experimental source: Japanese isolate
C;Comment: The cleavage sites of this polyprotein have not been determined.
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,234,250,305,325,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2

  Query Match             74.4%;  Score 11998.5;  DB 1;  Length 3010;
  Best Local Similarity   71.5%;  Pred. No. 0;
  Matches 2177;  Conservative  350;  Mismatches  472;  Indels   45;  Gaps    8;

Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
Db         61 RRQPIPKARRPEGRTWAQPGYPWPLYGNEGMGWAGWLLSPRGSRPSWGPTDPRRRSRNLG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
Db        121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
Db        181 LLSCLTIPASAYEVRNVSGIYHVTNDCSNSSIVYEAADMIMHTPGCVPCVRESNFSRCWV 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
Db        241 ALTPTLAARNSSIPTTTIRRHVDLLVGAAALCSAMYVGDLCGSVFLVSQLFTFSPRRYET 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
Db        301 VQDCNCSIYPGHVSGHRMAWDMMMNWSPTTALVVSQLLRIPQAVVDMVAGAHWGVLAGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
Db        361 YYSMVGNWAKVLIVMLLFAGVDGHTHVTGGRVASSTQSLVSWLSQGPSQKIQLVNTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
Db        421 HINRTALNCNDSLQTGFIAALFYAHRFNASGCPERMASCRPIDEFAQGWGPITHD--MPE 478

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
Db        479 SSDQRPYCWHYAPRPCGIVPASQVCGPVYCFTPSPVVVGTTDRFGAPTYSWGENETDVLL 538

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
Db        539 LSNTRPPQGNWFGCTWMNSTGFTKTCGGPPCNIGGVGN--NTLVCPTDCFRKHPEATYTK 596

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
Db        597 CGSGPWLTPRCMVDYPYRLWHYPCTVNFTVFKVRMYVGGVEHRLNAACNWTRGERCDLED 656

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
Db        657 RDRSELSPLLLSTTEWQILPCSFTTLPALSTGLIHLHRNIVDVQYLYGIGSAVVSFAIKW 716

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
Db        717 EYILLLFLLLADARVCACLWMMLLIAQAEATLENLVVLNAASVAGAHGLLSFLVFFCAAW 776

Qy        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
Db        777 YIKGRLVPGAAYALYGVWPLLLLLLALPPRAYAMDREMAASCGGAVFVGLVLLTLSPYYK 836

Qy        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
Db        837 VFLARLIWWLQYFITRAEAHLQVWVPPLNVRGGRDAIILLTCAVHPELIFDITKLLLAIL 896

Qy        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
Db        897 GPLMVLQAGITRVPYFVRAQGLIRACMLVRKVAGGHYVQMAFMKLAALTGTYVYDHLTPL 956

Qy        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
Db        957 RDWAHAGLRDLAVAVEPVVFSDMETKLITWGADTAACGDIISGLPVSARRGKEILLGPAD 1016

Qy       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080
Db       1017 SFGEQGWRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVDGEVQVLSTATQSFLATCVNG 1076

Qy       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
Db       1077 VCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWPAPPGARSMTPCTCGSSDLYLVTRH 1136

Qy       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200
Db       1137 ADVVPVRRRGDSRGSLLSPRPISYLKGSSGGPLLCPSGHVVGIFRAAVCTRGVAKAVDFI 1196

Qy       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260
Db       1197 PVESMETTMRSPVFTDNSSPPAVPQTFQVAHLHAPTGSGKSTKVPAAYAAQGYKVLVLNP 1256

Qy       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320
Db       1257 SVAATLGFGAYMSKAHGIEPNIRTGVRTIITGGPITYSTYCKFLADGGCSGGAYDIIICD 1316

Qy       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380
Db       1317 ECHSTDSTTILGIGTVLDQAETAGARLVVLATATPPGSITVPHPNIEEVALSNTGEIPFY 1376
```

```
Qy       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440
Db       1377 GKAIPIEAIKGGRHLIFCHSKKKCDELAAKLTGLGLNAVAYYRGLDVSVIPTSGDVVVVA 1436

Qy       1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500
Db       1437 TDALMTGFTGDFDSVIDCNTCVTQTVDFSLDPTFTIETTTLPQDAVSRAQRRGRTGRGRS 1496

Qy       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560
Db       1497 GIYRFVTPGERPSGMFDSSVLCECYDAGCAWYELTPAETSVRLRAYLNTPGLPVCQDHLE 1556

Qy       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
Db       1557 FWESVFTGLTHIDAHFLSQTKQAGDNLPYLVAYQATVCARAQAPPPSWDQMWKCLIRLKP 1616

Qy       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680
Db       1617 TLHGPTPLLYRLGAVQNEVTLTHPITKYIMACMSADLEVVTSTWVLVGGVLAALAAYCLT 1676

Qy       1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740
Db       1677 TGSVVIVGRIILSGRPAVIPDREVLYQEFDEMEECASHLPYIEQGMQLAEQFKQKALGLL 1736

Qy       1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
Db       1737 QTATKQAEAAAPVVESKWRALEVFWAKHMWNFISGIQYLAGLSTLPGNPAIASLMAFTAS 1796

Qy       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860
Db       1797 ITSPLTTQNTLLFNILGGWVAAQLAPPSAASAFVGAGIAGAAVGSIGLGKVLVDILAGYG 1856

Qy       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920
Db       1857 AGVAGALVAFKVMSGEMPSTEDLVNLLPAILSPGALVVGVVCAAILRRHVGPGEGAVQWM 1916

Qy       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980
Db       1917 NRLIAFASRGNHVSPTHYVPESDAAARVTQILSSLTITQLLKRLHQWINEDCSTPCSGSW 1976

Qy       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040
Db       1977 LKDVWDWICTVLSDFKTWLQSKLLPRLPGLPFLSCQRGYKGVWRGDGIMQTTCPCGAQIT 2036

Qy       2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100
Db       2037 GHVKNGSMRIVGPKTCSNTWHGTFPINAYTTGPCTPSPAPNYSRALWRVAAEEYVEVTRV 2096

Qy       2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160
Db       2097 GDFHYVTGMTTDNVKCPCQVPAPEFFTEVDGVRLHRYAPVCKPLLREEVVFQVGLNQYLV 2156

Qy       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220
Db       2157 GSQLPCEPEPDVAVLTSMLTDPSHITAETAKRRLARGSPPSLASSSASQLSAPSLKATCT 2216

Qy       2221 THGKAYDVDMVDANLF----MGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYML 2276
Db       2217 THHDSPDADLIEANLLWRQEMGGNITRVESENKVVILDSFDPIRAVEDEREISVPAEILR 2276

Qy       2277 PKKRFPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLS 2336
Db       2277 KPRKFPPALPIWARPDYNPPLLESWKDPDYVPPVVHGCPLPSTKAPPIPPPRRKRTVVLT 2336

Qy       2337 EDSIGDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQT-PPDELA-----LSETGSIS 2390
Db       2337 ESTVSSALAELATKTFG----------SSGSSAVDSGTATGPPDQASDDGDKGSDVESYS 2386

Qy       2391 SMPPLEGELGDPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCS-EEDDSVVCCSMSYSWT 2449
Db       2387 SMPPLEGEPGDPDL-------------------SDGSWSTVSGEAGEDVVCCSMSYTWT 2426

Qy       2450 GALITPCSPEEEKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSV 2509
Db       2427 GALITPCAAEESKLPINPLSNSLLRHHSMVYSTTSRSASLRQKKVTFDRLQVLDDHYRDV 2486

Qy       2510 LKDIKLAASKVTARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLL 2569
Db       2487 LKEMKAKASTVKARLLSIEEACKLTPPHSAKSKFGYGAKDVRSLSSRAVNHIRSVWEDLL 2546

Qy       2570 EDSETPIPTTIMAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVM 2629
Db       2547 EDTETPIDTTIMAKNEVFCVQPEKGGRKPARLIVFPDLGVRVCEKMALYDVVSTLPQAVM 2606

Qy       2630 GASYGFQYSPAQRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPE 2689
Db       2607 GPSYGFQYSPGQRVEFLVNTWKSKKCPMGFSYDTRCFDSTVTENDIRTEESIYQCCDLAP 2666

Qy       2690 EAHTAIHSLTERLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAG 2749
Db       2667 EARQAIRSLTERLYVGGPLTNSKGQNCGYRRCRASGVLTTSCGNTLTCYLKATAACRAAK 2726

Qy       2750 IIAPTMLVCGDDLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCS 2809
Db       2727 LQDCTMLVNGDDLVVICESAGTQEDAAALRAFTEAMTRYSAPPGDPPQPEYDLELITSCS 2786

Qy       2810 SNVSVALGPQGRRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMT 2869
Db       2787 SNVSVAHDASGKRVYYLTRDPTTPLARAAWETVRHTPVNSWLGNIIMYAPTLWARMILMT 2846

Qy       2870 HFFSILMAQDTLDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVAS 2929
Db       2847 HFFSILLAQEQLEKALDCQIYGACYSIEPLDLPQIIERLHGLSAFSLHSYSPGEINRVAS 2906

Qy       2930 ALRKLGAPPLRAWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDL 2989
Db       2907 CLRKLGVPPLRVWRHRARSVRAKLLSQGGRAATCGKYLFNWAVKTKLKLTPIPAASQLDL 2966

Qy       2990 SSWFTVGAGGGDIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
Db       2967 SGWFVAGYNGGDIYHSLSRARPRWFMLCLLLLSVGVGIYLLPNR 3010
```

RESULT 4
GNWVTC
genome polyprotein - hepatitis C virus
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 09-Jul-2004
C;Accession: A38465
R;Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E.;
J. Virol. 65, 1105-1113, 1991
A;Title: Structure and organization of the hepatitis C virus genome isolated from human
A;Reference number: A38465; MUID:91140698; PMID:1847440
A;Accession: A38465
A;Molecule type: genomic RNA
A;Residues: 1-3010 <TAK>
A;Cross-references: UNIPROT:P26663; EMBL:M58335; NID:g329770; PIDN:AAA72945.1; PID:g3297
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,22

Query Match 74.3%; Score 11984.5; DB 1; Length 3010;

```
  Best Local Similarity   71.5%;  Pred. No. 0;
  Matches 2175;  Conservative  352;  Mismatches  472;  Indels   45;  Gaps    8;

Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRAPRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              ||||||| ||   |::| :|||||||||||||||||||||||||||||| ||| ||||:|
Db         61 RRQPIPKARRPEGRTWAQPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPTDPRRRSRNLG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
              |||||||||||||||||:|||||||  ||||||||||||||||:|||||||||||||||
Db        121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
              ||||:||| || || |:|  | |||||:| || ::    ::| |||||| : ||:|:||:
Db        181 LLSCLTTPASAYEVHNVSGIYHVTNDCSNASIVYEAADLIMHTPGCVPCVREGNSSRCWV 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
              ::| :| :     |  :| |:|::| :|  |||:||||||| | | :|:|  ||: |
Db        241 ALTPTLAARNVTIPTTTIRRHVDLLVGAAAFCSAMYVGDLCGSVFLVSQLFTFSPRRHVT 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
              :|||||||||| ::|||||||||||||||  ::::  :|:|: ::|:::||||||: |||
Db        301 LQDCNCSIYPGHVSGHRMAWDMMMNWSPTTALVVSQLLRIPQAVVDMVAGAHWGVLAGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
              |:|| | ||||::::|| ||||  ||  ||: |:|| || |:|  || |||||:||||||
Db        361 YYSMAGNWAKVLIVMLLFAGVDGDTHVTGGAQAKTTNRLVSMFASGPSQKIQLINTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
              |||||||||||| |||:|:|||||||||||||||: ||:|: |  ||| : | :  ::
Db        421 HINRTALNCNDSLQTGFLAALFYTHSFNSSGCPERMAQCRTIDKFDQGWGPITYAE--SS 478

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
              | |||||||||  || :| |  ||||||||||||||||||| | ||| ||||||||  |
Db        479 RSDQRPYCWHYPPPQCTIVPASEVCGPVYCFTPSPVVVGTTDRFGVPTYRWGENETDVLL 538

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
              ||:|||| |:|||||||:|:||||| |||       |    | ||||||||||: || |
Db        539 LNNTRPPQGNWFGCTWMNSTGFTKTCGGPPCNIGGVGN--NTLTCPTDCFRKHPEATYTK 596

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
              ||||||||||::|||||||||||||:||||:|||||||||| ||||:|||:||:|||
Db        597 CGSGPWLTPRCMVDYPYRLWHYPCTVNFTIFKVRMYVGGVEHRLNAACNWTRGERCDLED 656

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
              ||| :|||| |||| :|||:: |||||||:||||||||::||:  |:  : ::|
Db        657 RDRPELSPLLLSTTEWQVLPCSFTTLPALSTGLIHLHQNIVDVQYLYGIGSAVVSFAIKW 716

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
              |:|:|||||||||||||::|: ||||||| ||:|::|| |  :| | |::|| |||
Db        717 EYVLLLFLLLADARVCACLWMMLLIAQAEAALENLVVLNSASVAGAHGILSFLVFFCAAW 776

Qy        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
              |||||:|| |||:| |:|  |||||| :||| |  :   | |: | : | ||:| ||
Db        777 YIKGRLVPGATYALYGVWPLLLLLLALPPRAYAMDREMAASCGGAVFVGLVLLTLSPYYK 836

Qy        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
              |:| :||| |  |  || :  | ||:  ||||| ||  :  :| ::||||| |:|:|
Db        837 VFLARLIWWLQYFTTRAEADLHVWIPPLNARGGRDAIILLMCAVHPELIFDITKLLIAIL 896

Qy        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
              ||  :|:  :|||||||||  |:  | : | :||| ||||| : ||  |||||:||||:
Db        897 GPLMVLQAGITRVPYFVRAQGLIHACMLVRKVAGGHYVQMAFMKLGALTGTYIYNHLTPL 956

Qy        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
              ||  :|||||||||::|| || |:| |||:|||||||: ||||||| |:|:||||||
Db        957 RDWPRAGLRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGKEILLGPAD 1016

Qy       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080
                :|  ||||||||:|||||||| |: |:|||||  :  ||:||:|| ||||| | ::|

Db       1017 SLEGRGLRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNG 1076

Qy       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
              | ||||||||:|||| :||:||||:: : |||||| ||| :|| |||||: |||||||:
Db       1077 VCWTVYHGAGSKTLAAPKGPITQMYTNVDQDLVGWPKPPGARSLTPCTCGSSDLYLVTRH 1136

Qy       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200
              ||||| ||||| ||:|||||:| |||||||||:||| |||||:||||||:|||||::||:
Db       1137 ADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPFGHAVGIFRAAVCTRGVAKAVDFV 1196

Qy       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260
              |||:::   ||| |:|||:||||||::|| :|||||||||||||| ||||||||||||||
Db       1197 PVESMETTMRSPVFTDNSSPPAVPQSFQVAHLHAPTGSGKSTKVPAAYAAQGYKVLVLNP 1256

Qy       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320
              ||||||||||:||||||:|||||||||:|||||:|||||||||||||:||||||||||
Db       1257 SVAATLGFGAYMSKAHGIDPNIRTGVRTITTGAPVTYSTYGKFLADGGCSGGAYDIIICD 1316

Qy       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380
              |||: |||||||||||||||||| || |||||||||||| ||||||||||   ||||||
Db       1317 ECHSTDSTTILGIGTVLDQAETAGARLVVLATATPPGSVTVPHPNIEEVALSNTGEIPFY 1376

Qy       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440
              |:|||:  |:|||||||||||||||||| | |:|:|:|||||||||||||| |||||||
Db       1377 GKAIPIEAIRGGRHLIFCHSKKKCDELAAKLSGLGINAVAYYRGLDVSVIPTIGDVVVVA 1436

Qy       1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500
              |||||||||||||||||  ||| |||||||||| |  |||||||||||||||||||
Db       1437 TDALMTGYTGDFDSVIDCNTCVTQTVDFSLDPTFTIETTTVPQDAVSRSQRRGRTGRGRR 1496

Qy       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560
              ||||:|: ||| |||||| |||||||:||:|||||| ||||||||||||||||||||||
Db       1497 GIYRFVTPGERPSGMFDSSVLCECYDAGCAWYELTPAETSVRLRAYLNTPGLPVCQDHLE 1556

Qy       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
              |||:|||||||||||||||:|:|| || |||||||||:||||||| ||||| ||||
Db       1557 FWESVFTGLTHIDAHFLSQTKQAGDNFPYLVAYQATVCARAQAPPPSWDQMWKCLIRLKP 1616

Qy       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680
              || |||||||||:| |||||||:|||| || |||||:|||||| ||||||:|||||
Db       1617 TLHGPTPLLYRLGAVQNEVTLTHPITKYIMACMSADLEVVTSTWVLVGGVLAALAAYCLT 1676

Qy       1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740
              || | |:||: :: |  : ||:|:||: |||||||||     ||:| ::||  | |  |||
Db       1677 TGSVVIVGRIILSGRPAIVPDRELLYQEFDEMEECASHLPYIEQGMQLAEQFKQKALGLL 1736

Qy       1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
              | |:|||:   |  |:: |  :| |||||||||||||||||||||||||:||:|||:|:
Db       1737 QTATKQAEAAAPVVESKWRALETFWAKHMWNFISGIQYLAGLSTLPGNPAIASLMAFTAS 1796

Qy       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860
              :||||:| :|:| ||||||:|:|:|||: |: || :|: ||||||||||||||||||||
Db       1797 ITSPLTTQSTLLFNILGGWVAAQLAPPSAASAFVGAGIAGAAVGSIGLGKVLVDILAGYG 1856

Qy       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920
              ||::|||||:|||| || ||:|||| ||||||||||:||||||||||||||||||||||
Db       1857 AGVAGALVAFKVMSGEMPSTEDLVNLLPAILSPGALVVGVVCAAILRRHVGPGEGAVQWM 1916

Qy       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980
              |||||||||||:|||| |||| ||||:| |||| ||:||| || |||  || |||
Db       1917 NRLIAFASRGNHVSPTHYVPESDAAARVTQILSSLTITQLLKRLHQWINEDCSTPCSGSW 1976

Qy       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040
              |||||||:||:|||| || ||| |::||:|| |||:|||||| | ||| | |||| |:
Db       1977 LRDVWDWICTVLTDFKTWLQSKLLPQLPGVPFFSCQRGYKGVWRGDGIMQTTCPCGAQIT 2036

Qy       2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100
              |:|: ||||| ||||| | | |||||| || | | | ||||:  |:||||| || |||:
Db       2037 GHVKNGSMRIVGPKTCSNTWHGTFPINAYTTGPCTPSPAPNYSRALWRVAAEEYVEVTRV 2096

Qy       2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160
              | :||:||:||||:| |||:|:||||| ||||::||:||  :|  |:||:| |||| ::|
Db       2097 GDFHYVTGMTTDNVKCPCQVPAPEFFSEVDGVRLHRYAPACRPLLREEVTFQVGLNQYLV 2156
```

```
Qy       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220
Db       2157 GSQLPCEPEPDVAVLTSMLTDPSHITAETAKRRLARGSPPSLASSSASQLSAPSLKATCT 2216

Qy       2221 THGKAYDVDMVDANLF----MGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYML 2276
Db       2217 THHVSPDADLIEANLLWRQEMGGNITRVESENKVVVLDSFDPLRAEEDEREVSVPAEILR 2276

Qy       2277 PKKRFPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLS 2336
Db       2277 KSKKFPAAMPIWARPDYNPPLLESWKDPDYVPPVVHGCPLPPIKAPPIPPPRRKRTVVLT 2336

Qy       2337 EDSIGDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTP-PDELA-----LSETGSIS 2390
Db       2337 ESSVSSALAELATKTFG----------SSESSAVDSGTATALPDQASDDGDKGSDVESYS 2386

Qy       2391 SMPPLEGELGDPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEE-DDSVVCCSMSYSWT 2449
Db       2387 SMPPLEGEPGDPDL-------------------SDGSWSTVSEEASEDVVCCSMSYTWT 2426

Qy       2450 GALITPCSPEEEKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSV 2509
Db       2427 GALITPCAAEESKLPINALSNSLLRHHNMVYATTSRSAGLRQKKVTFDRLQVLDDHYRDV 2486

Qy       2510 LKDIKLAASKVTARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLL 2569
Db       2487 LKEMKAKASTVKAKLLSVEEACKLTPPHSAKSKFGYGAKDVRNLSSKAVNHIHSVWKDLL 2546

Qy       2570 EDSETPIPTTIMAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVM 2629
Db       2547 EDTVTPIDTTIMAKNEVFCVQPEKGGRKPARLIVFPDLGVRVCEKMALYDVVSTLPQVVM 2606

Qy       2630 GASYGFQYSPAQRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPE 2689
Db       2607 GSSYGFQYSPGQRVEFLVNTWKSKKNPMGFSYDTRCFDSTVTENDIRVEESIYQCCDLAP 2666

Qy       2690 EAHTAIHSLTERLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAG 2749
Db       2667 EARQAIKSLTERLYIGGPLTNSKGQNCGYRRCRASGVLTTSCGNTLTCYLKASAACRAAK 2726

Qy       2750 IIAPTMLVCGDDLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCS 2809
Db       2727 LQDCTMLVNGDDLVVICESAGTQEDAASLRVFTEAMTRYSAPPGDPPQPEYDLELITSCS 2786

Qy       2810 SNVSVALGPQGRRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMT 2869
Db       2787 SNVSVAHDASGKRVYYLTRDPTTPLARAAWETARHTPVNSWLGNIIMYAPTLWARMILMT 2846

Qy       2870 HFFSILMAQDTLDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVAS 2929
Db       2847 HFFSILLAQEQLEKALDCQIYGACYSIEPLDLPQIIERLHGLSAFSLHSYSPGEINRVAS 2906

Qy       2930 ALRKLGAPPLRAWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDL 2989
Db       2907 CLRKLGVPPLRVWRHRARSVRARLLSQGGRAATCGKYLFNWAVKTKLKLTPIPAASRLDL 2966

Qy       2990 SSWFTVGAGGGDIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
Db       2967 SGWFVAGYSGGDIYHSLSRARPRWFMLCLLLLSVGVGIYLLPNR 3010


RESULT 5
A45573
genome polyprotein - hepatitis C virus (strain JT)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 09-Jul-2004
C;Accession: A45573
R;Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata,
Virus Res. 23, 39-53, 1992
A;Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier: s
A;Reference number: A45573; MUID:92295714; PMID:1318627
A;Accession: A45573
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-3010 <TAN>
A;Cross-references: UNIPROT:Q00269; GB:D11168; GB:D01171; NID:g221612; PIDN:BAA01943.1;
A;Experimental source: HCV-JT
A;Note: sequence extracted from NCBI backbone (NCBIN:106206, NCBIP:106207)
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

  Query Match             74.3%;  Score 11979.5;  DB 1;  Length 3010;
  Best Local Similarity   71.3%;  Pred. No. 0;
  Matches 2172;  Conservative  358;  Mismatches  467;  Indels   49;  Gaps    9;

Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
Db          1 MSTNPKPQRKTKRNTYRRPQDVKFPGGGQIVGGVYVLPRRGPTLGVRATRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
Db         61 RRQPIPKARRPEGRAWAQPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPTDPRRRSRNLG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
Db        121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
Db        181 LLSCLTIPASAYQVRNASGLYHVTNDCSNSSIVYEAAGMIMHTPGCVPCVRENNASRCWV 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
Db        241 ALTPTLAARNTSIPTTTIRRHVDLLVGAAAFCSAMYVGDLCGSVFLVSQLFTFSPRRYET 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
Db        301 VQDCNCSIYPGHVSGHRMAWDMMMNWSPTTALVVSQLLRIPQAVVDMVAGAHWGVLAGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
Db        361 YYSMVGNWAKVLIVMLLFAGVDGVTYTTGGSQARHTQSVTSFFTQGPAQRIQLINTNGSW 420

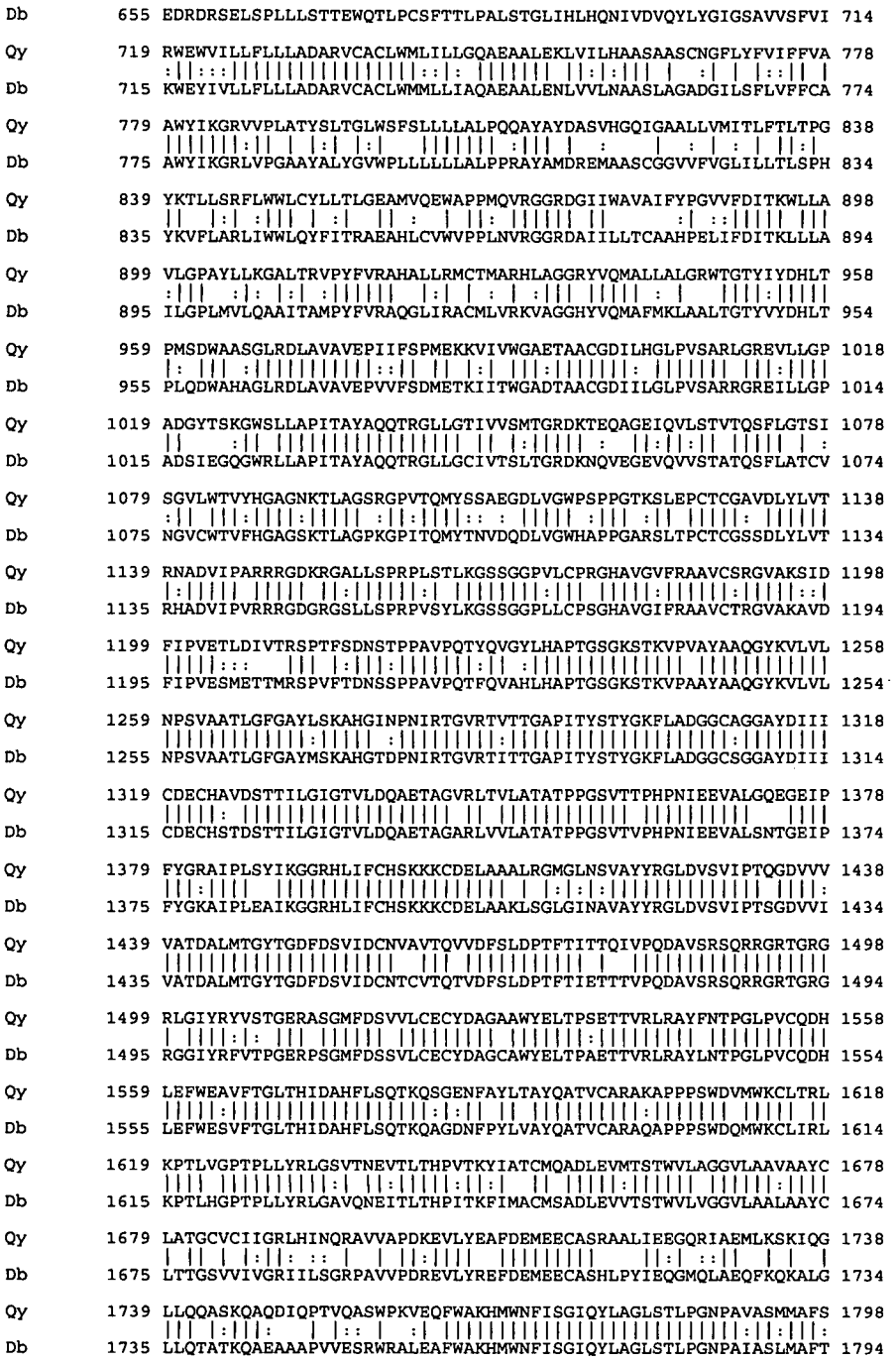
Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
Db        421 HINRTALNCNESLNTGFFAALFYAHKFNSSGCPERMASCSSIDKFAQGWGPITY----TE 476

Qy        481 PEDM--RPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDV 538
Db        477 PRDLDQRPYCWHYAPRQCGIVPASQVCGPVYCFTPSPVVVGTTDRSGAPTYNWGANETDV 536

Qy        539 FLLNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTY 598
Db        537 LLLNNTRPPQGNWFGCTWMNSTGFTKTCGGPPCNIGGVGNLT--LTCPTDCFRKHPEATY 594

Qy        599 LKCGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNL 658
Db        595 TKCGSGPWLTPRCIVDYPYRLWHYPCTVNFTIFKVRMYVGGVEHRLSAACNWTRGERCDL 654

Qy        659 EDRDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIV 718
```


Db 655 EDRDRSELSPLLLSTTEWQTLPCSFTTLPALSTGLIHLHQNIVDVQYLYGIGSAVVSFVI 714
Qy 719 RWEWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVA 778
Db 715 KWEYIVLLFLLLADARVCACLWMMLLIAQAEAALENLVVLNAASLAGADGILSFLVFFCA 774
Qy 779 AWYIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPG 838
Db 775 AWYIKGRLVPGAAYALYGVWPLLLLLLALPPRAYAMDREMAASCGGVVFVGLILLTLSPH 834
Qy 839 YKTLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLA 898
Db 835 YKVFLARLIWWLQYFITRAEAHLCVWVPPLNVRGGRDAIILLTCAAHPELIFDITKLLLA 894
Qy 899 VLGPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLT 958
Db 895 ILGPLMVLQAAITAMPYFVRAQGLIRACMLVRKVAGGHYVQMAFMKLAALTGTYVYDHLT 954
Qy 959 PMSDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGP 1018
Db 955 PLQDWAHAGLRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGREILLGP 1014
Qy 1019 ADGYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSI 1078
Db 1015 ADSIEGQGWRLLAPITAYAQQTRGLLGCIVTSLTGRDKNQVEGEVQVVSTATQSFLATCV 1074
Qy 1079 SGVLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVT 1138
Db 1075 NGVCWTVFHGAGSKTLAGPKGPITQMYTNVDQDLVGWHAPPGARSLTPCTCGSSDLYLVT 1134
Qy 1139 RNADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSID 1198
Db 1135 RHADVIPVRRRGDGRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVD 1194
Qy 1199 FIPVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVL 1258
Db 1195 FIPVESMETTMRSPVFTDNSSPPAVPQTFQVAHLHAPTGSGKSTKVPAAYAAQGYKVLVL 1254
Qy 1259 NPSVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIII 1318
Db 1255 NPSVAATLGFGAYMSKAHGTDPNIRTGVRTITTGAPITYSTYGKFLADGGCSGGAYDIII 1314
Qy 1319 CDECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIP 1378
Db 1315 CDECHSTDSTTILGIGTVLDQAETAGARLVVLATATPPGSVTVPHPNIEEVALSNTGEIP 1374
Qy 1379 FYGRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVV 1438
Db 1375 FYGKAIPLEAIKGGRHLIFCHSKKKCDELAAKLSGLGINAVAYYRGLDVSVIPTSGDVVI 1434
Qy 1439 VATDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRG 1498
Db 1435 VATDALMTGYTGDFDSVIDCNTCVTQTVDFSLDPTFTIETTTVPQDAVSRSQRRGRTGRG 1494
Qy 1499 RLGIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDH 1558
Db 1495 RGGIYRFVTPGERPSGMFDSSVLCECYDAGCAWYELTPAETTVRLRAYLNTPGLPVCQDH 1554
Qy 1559 LEFWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRL 1618
Db 1555 LEFWESVFTGLTHIDAHFLSQTKQAGDNFPYLVAYQATVCARAQAPPPSWDQMWKCLIRL 1614
Qy 1619 KPTLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYC 1678
Db 1615 KPTLHGPTPLLYRLGAVQNEITLTHPITKFIMACMSADLEVVTSTWVLVGGVLAALAAYC 1674
Qy 1679 LATGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQG 1738
Db 1675 LTTGSVVIVGRIILSGRPAVVPDREVLYREFDEMEECASHLPYIEQGMQLAEQFKQKALG 1734
Qy 1739 LLQQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFS 1798
Db 1735 LLQTATKQAEAAAPVVESRWRALEAFWAKHMWNFISGIQYLAGLSTLPGNPAIASLMAFT 1794
Qy 1799 AALTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAG 1858
Db 1795 ASITSPLTTQNTLLFNILGGWVAAQLAPPSAASAFVGAGIAGAAIGSIGLGKVLVDILAG 1854
Qy 1859 YGAGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQ 1918
Db 1855 YGAGVAGALVAFKVMSGEAPSAEDLVNLLPAILSPGALVVGVVCAAILRRHVGPGEGAVQ 1914
Qy 1919 WMNRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGG 1978
Db 1915 WMNRLIAFASRGNHVSPTHYVPESDAAARVTQILSSLTITQLLKRLHQWINEDCSTPCSG 1974
Qy 1979 SWLRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGAN 2038
Db 1975 SWLKDVWDWICTVLTDFKTWLQSKLLPKLPGVPFFSCQRGYKGVWRGDGIMQTTCPCGAQ 2034
Qy 2039 ISGNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVT 2098
Db 2035 ITGHVKNGSMRIVGPKTCSNTWHGTFPINAYTTGPCTPSPAPNYSRALWRVAAEEYVEIT 2094
Qy 2099 QHGSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSF 2158
Db 2095 RVGDFHYVTGMTTDNVKCPCQVPAPEFFTELDGVRLHRYAPACRPLLREDVTFQVGLNQY 2154
Qy 2159 VVGSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRAT 2218
Db 2155 LVGSQLPCEPEPDVAVLTSMLTDPSHITAETAKRRLARGSPPSLASSSASQLSAPSLKAT 2214
Qy 2219 CTTHGKAYDVDMVDANLF----MGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEY 2274
Db 2215 CTTHHDSPDADLIEANLLWRQEMGGNITRVESENKVVILDSFDPLRAEEDEREVSVAAEI 2274
Qy 2275 MLPKKRFPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVG 2334
Db 2275 LRKSKKFPPALPIWARPDYNPPLLESWKSPDYVPPAVHGCPLPPTTGPPIPPPRKKRTVV 2334
Qy 2335 LSEDSIGDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQT-PPDELA-----LSETGS 2388
Db 2335 LTESTVSSALAELATKTFG----------SSGSSAVDSGTATAPPDQTSDDGDKESDVES 2384
Qy 2389 ISSMPPLEGELGDPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCS-EEDDSVVCCSMSYS 2447
Db 2385 YSSMPPLEGEPGDPDL--------------------SDGSWSTVSGEASDDIVCCSMSYT 2424
Qy 2448 WTGALITPCSPEEEKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYD 2507
Db 2425 WTGALITPCAAEESKLPINALSNSLLRHHNMVYATTSRSASLRQKKVTFDRLQVLDDHYR 2484
Qy 2508 SVLKDIKLAASKVTARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKD 2567
Db 2485 DVLKEMKAKASTVKAKLLSVEEACKLTPPHSAKSKFGYGAKDVRNLSSKAINHIRSVWKD 2544
Qy 2568 LLEDSETPIPTTIMAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQA 2627
Db 2545 LLEDTETPIDTTIMAKSEVFCVQPEKGGRKPARLIVFPDLGVRVCEKMALYDVVSTLPQA 2604
Qy 2628 VMGASYGFQYSPAQRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSL 2687
Db 2605 VMGSSYGFQYSPGQRVEFLVNAWKSKKSPMGFSYDTRCFDSTVTESDIRVEESIYQCCDL 2664
Qy 2688 PEEAHTAIHSLTERLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKA 2747
Db 2665 APEARQAIKSLTERLYIGGPLTNSKGQNCGYRRCRASGVLTTSCGNTLTCYLKATAACRA 2724
Qy 2748 AGIIAPTMLVCGDDLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITS 2807
Db 2725 AKLQDCTMLVNGDDLVVICESAGTQEDAASLRVFTEAMTRYSAPPGDPPQPEYDLELITS 2784
Qy 2808 CSSNVSVALGPQGRRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVL 2867
Db 2785 CSSNVSVAHDASGKRVYYLTRDPTTPLARAAWETARHTPVNSWLGNIIMYAPTLWARMIL 2844

```
Qy       2868 MTHFFSILMAQDTLDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRV 2927
Db       2845 MTHFFSILLAQEQLEKALDCQIYGACYSIEPLDLPQIIQRLHGLSAFSLHSYSPGEINRV 2904

Qy       2928 ASALRKLGAPPLRAWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLL 2987
Db       2905 ASCLRKLGVPPLRVWRHRARSVRARLLSQGGRAATCGKYLFNWAVRTKLKLTPIPAASQL 2964

Qy       2988 DLSSWFTVGAGGGDIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
Db       2965 DLSSWFVAGYSGGDIYHSLSRARPRWFMWCLLLLSVGVGIYLLPNR 3010


RESULT 6
GNWVC3
genome polyprotein - hepatitis C virus (strain HCV-1)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 09-Jul-2004
C;Accession: A39166; PQ0403; PQ0404
R;Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Coi
Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991
A;Title: Genetic organization and diversity of the hepatitis C virus.
A;Reference number: A39166; MUID:91172826; PMID:1848704
A;Accession: A39166
A;Molecule type: mRNA
A;Residues: 1-3011 <CHO>
A;Cross-references: UNIPROT:P26664; GB:M62321; NID:g329873; PIDN:AAA45676.1; PID:g329874
R;Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.L
J. Gen. Virol. 73, 1131-1141, 1992
A;Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e
A;Reference number: PQ0393; MUID:92268871; PMID:1316939
A;Accession: PQ0403
A;Molecule type: genomic RNA
A;Residues: 1577-1633 <CHA>
A;Cross-references: DDBJ:D10128
A;Experimental source: isolates E-b16
A;Accession: PQ0404
A;Status: preliminary
A;Molecule type: genomic RNA
A;Residues: 1577-1633 <CH2>
A;Experimental source: isolates E-b17
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
F;1-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077,22

  Query Match             74.0%;  Score 11922;  DB 1;  Length 3011;
  Best Local Similarity   71.4%;  Pred. No. 0;
  Matches 2172;  Conservative  352;  Mismatches  478;  Indels   40;  Gaps   12;

Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
Db          1 MSTNPKPQKKNKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
Db         61 RRQPIPKARRPEGRTWAQPGYPWPLYGNEGCGWAGWLLSPRGSRPSWGPTDPRRRSRNLG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
Db        121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTG-YMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCW 239
Db        181 LLSCLTVPASAYQVRN-STGLYHVTNDCPNSSIVYEAADAILHTPGCVPCVREGNASRCW 239

Qy        240 IPVSPNVAVQRPGAL--TQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQH 297
Db        240 VAMTPTVAT-RDGKLPATQ-LRRHIDLLVGSATLCSALYVGDLCGSVFLVGQLFTFSPRR 297

Qy        298 HWFVQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMF 357
Db        298 HWTTQGCNCSIYPGHITGHRMAWDMMMNWSPTTALVMAQLLRIPQAILDMIAGAHWGVLA 357

Qy        358 GLAYFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTN 417
Db        358 GIAYFSMVGNWAKVLVVLLLFAGVDAETHVTGGSAGHTVSGFVSLLAPGAKQNVQLINTN 417

Qy        418 GSWHINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDN 477
Db        418 GSWHLNSTALNCNDSLNTGWLAGLFYHHKFNSSGCPERLASCRPLTDFDQGWGPISYA-N 476

Qy        478 VTNPEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETD 537
Db        477 GSGP-DQRPYCWHYPPKPCGIVPAKSVCGPVYCFTPSPVVVGTTDRSGAPTYSWGENDTD 535

Qy        538 VFLLNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTT 597
Db        536 VFVLNNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVIGGAGN--NTLHCPTDCFRKHPDAT 593

Qy        598 YLKCGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCN 657
Db        594 YSRCGSGPWITPRCLVDYPYRLWHYPCTINYTIPKIRMYVGGVEHRLEAACNWTRGERCD 653

Qy        658 LEDRDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYI 717
Db        654 LEDRDRSELSPLLLTTTQWQVLPCSFTTLPALSTGLIHLHQNIVDVQYLYGVGSSIASWA 713

Qy        718 VRWEWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFV 777
Db        714 IKWEYVVLLFLLLADARVCSCLWMMLLISQAEAALENLVILNAASLAGTHGLVSFLVFFC 773

Qy        778 AAWYIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTP 837
Db        774 FAWYLKGKWVPGAVYTFYGMWPLLLLLLALPQRAYALDTEVAASCGGVVLVGLMALTLSP 833

Qy        838 GYKTLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLL 897
Db        834 YYKRYISWCLWWLQYFLTRVEAQLHVWIPPLNVRGGRDAVILLMCAVHPTLVFDITKLLL 893

Qy        898 AVLGPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHL 957
Db        894 AVFGPLWILQASLLKVPYFVRVQGLLRFCALARKMIGGHYVQMVIIKLGALTGTYVYNHL 953

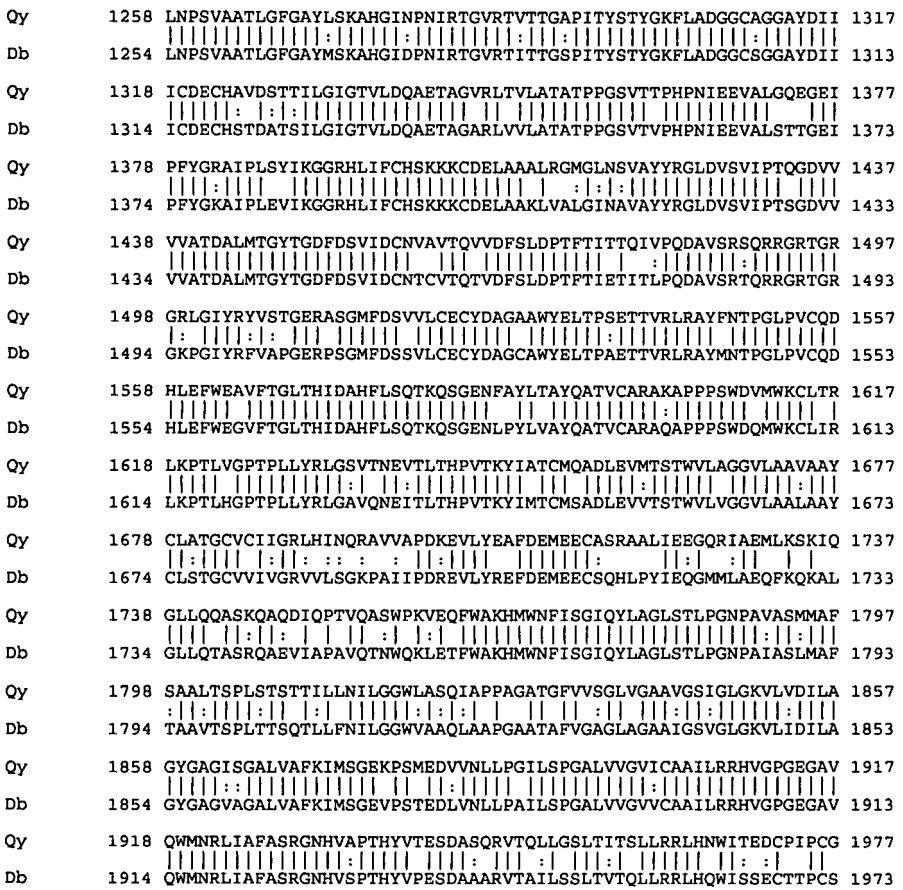
Qy        958 TPMSDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLG 1017
Db        954 TPLRDWAHNGLRDLAVAVEPVVFSQMETKLITWGADTAACGDIINGLPVSARRGREILLG 1013

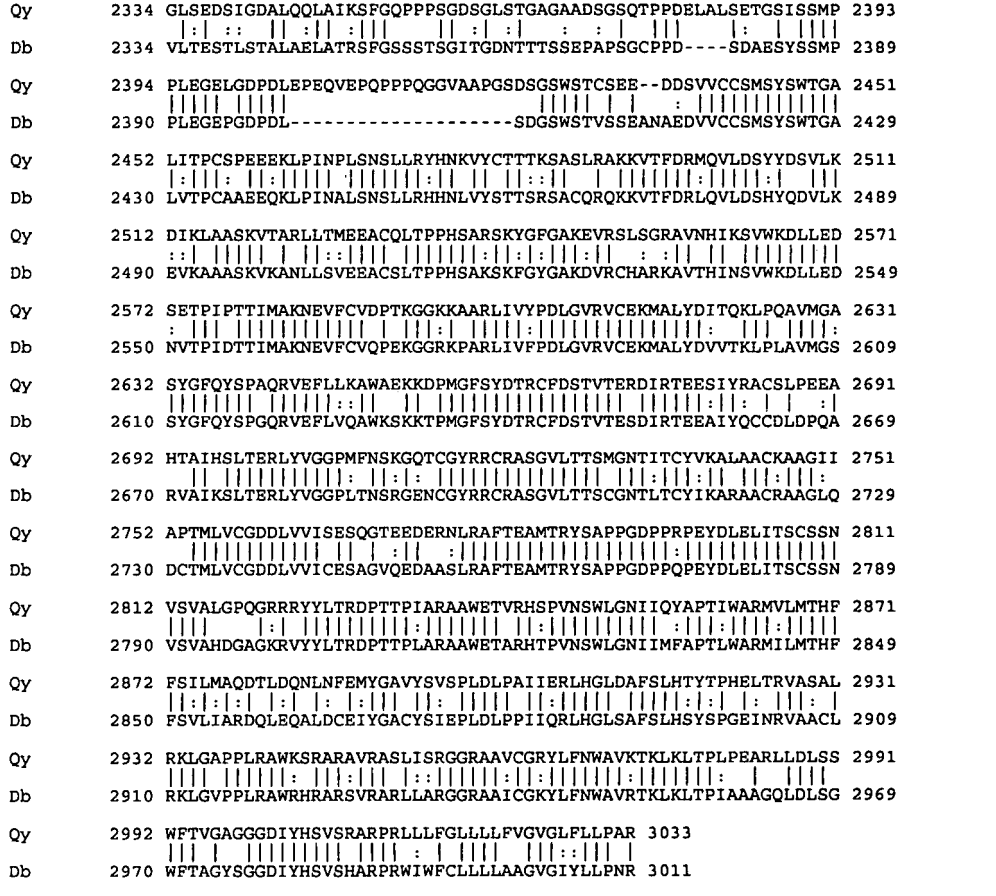
Qy       1018 PADGYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTS 1077
Db       1014 PADGMVSKGWRLLAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATC 1073

Qy       1078 ISGVLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLV 1137
Db       1074 INGVCWTVYHGAGTRTIASPKGPVIQMYTNVDQDLVGWPAPQGSRSLTPCTCGSSDLYLV 1133

Qy       1138 TRNADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSI 1197
Db       1134 TRHADVIPVRRRGDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAV 1193

Qy       1198 DFIPVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLV 1257
Db       1194 DFIPVENLETTMRSPVFTDNSSPPVVPQSFQVAHLHAPTGSGKSTKVPAAYAAQGYKVLV 1253
```





RESULT 7
GNWVTW
genome polyprotein - hepatitis C virus (strain Taiwan)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
A;Note: host Homo sapiens (man)
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 09-Jul-2004
C;Accession: A40244
R;Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.
Virology 188, 102-113, 1992
A;Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the
A;Reference number: A40244; MUID:92230206; PMID:1314449
A;Accession: A40244
A;Molecule type: genomic RNA
A;Residues: 1-3010 <CHE>
A;Cross-references: UNIPROT:P29846; GB:M84754
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
F;1-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

```
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,233,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,207

  Query Match             73.9%;  Score 11911.5;  DB 1;  Length 3010;
  Best Local Similarity   71.1%;  Pred. No. 0;
  Matches 2163;  Conservative  352;  Mismatches  484;  Indels   45;  Gaps    8;

Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              |||| |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db          1 MSTNGKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTWERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              ||||||| |:  |::| :|||||||||||||||||||:||||||:||| ||| ||||:|
Db         61 RRQPIPKARQPEGRAWAQPGYPWPLYGNEGLGWAGWLVSPRGSRPNWGPTDPRRRSRNLG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
              ||||||||||||||||||:||||||||||||||||||||:||||||||||||||||||
Db        121 KVIDTLTCGFADLMGYIPLVGAPLGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
              ||||:| | || || |:|  | |||||:| || ::    ::| |||||| :  |:|:||:
Db        181 LLSCLTIPASAYEVHNVSGIYHVTNDCSNSSIVYEAADMIMHTPGCVPCVRENNSSRCWV 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
               ::| :| :     |  :| |:|::| :|  |||:||||||| | | :|:|  ||: :
Db        241 ALTPTLAAARNNSVPTATIRRHVDLLVGAAAFCSAMYVGDLCGSVFLVSQLFTFSPRRYET 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
              |||||||||| :|||||||||||||||  ::::  :|:|: ::|:: ||||||: |||
Db        301 VQDCNCSIYPGHVTGHRMAWDMMMNWSPTTALVVSQLLRIPQAVVDMVGGAHWGVLAGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
              |:|| | ||||::::|| |||| |   ||: |:||  | |||  |  |||||:||||||
Db        361 YYSMVGNWAKVLIVMLLFAGVDGSTIVSGGTVARTTHSLASLFTQGASQKIQLINTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
              |||||||||||| |||:||||| | ||:|||||||::||||: |  ||| : | :   :
Db        421 HINRTALNCNDSLQTGFLASLFYAHRFNASGCPERMASCRSIDKFDQGWGPITYTE--AD 478

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
              :| ||||||| || ||:| |  ||||||||||||||||||| |||||:|||||||  :
Db        479 IQDQRPYCWHYAPRPCGIVPASQVCGPVYCFTPSPVVVGTTDRFGAPTYSWGENETDVLI 538

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
              ||:|||| |:|||||||||:|:|||||| |||       |    |:||||||||: || |
Db        539 LNNTRPPQGNWFGCTWMNSTGFTKTCGGPPCNIGGGGN--NTLVCPTDCFRKHPEATYTK 596

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
              |||||||||||:|||||||||||||:||||:|||||||||| ||||:|||:||:|||
Db        597 CGSGPWLTPRCMVDYPYRLWHYPCTVNFTIFKVRMYVGGVEHRLNAACNWTRGERCDLED 656

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
              ||||:||||| ||||| |||||:: |||||||:||||:||||::||:  |:  : ::|
Db        657 RDRSELSPLLLSTTEWQILPCSFTGLPALSTGLIHLHQNVVDVQYLYGIGSAVVSFAIKW 716

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
              |:::|||||||||||||||::|: ||||||| ||: :||| |  :|  | |::|| |||
Db        717 EYILLLFLLLADARVCACLWMMLLIAQAEAALENLVVFNAASVAGMHGTLSFLVFFCAAW 776

Qy        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
              |||||:|| | |:|  |:|   ||||||| :||| |  :   | |: | : | ||:| ||
Db        777 YIKGRLVPGAAYALYGVWPLLLLLLALPPRAYAMDREMAASCGGAVFVGLVLLTLSPHYK 836

Qy        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
              |:| :||| | :|  || :| | ||: |||||  ||      || ::||||| |||:|
Db        837 MFLARLIWWLQYFITRAEAHLQVWIPPLNVRGGRDAIILLTCAAYPELIFDITKILLAIL 896

Qy        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
              || :|:  |||:|||||  |:| | : |  ||| |||||: |   ||||:||||||:
Db        897 GPLMVLQAGLTRIPYFVRAQGLIRACMLVRKAAGGHYVQMALMKLAALTGTYVYDHLTPL 956

Qy        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
              ||| :|||||||||::|| || |:| |||:|||||: |||||| |||:||||||
Db        957 QDWAHTGLRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGREILLGPAD 1016

Qy       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080
                 :|| |||||||||||||| | |: |:|||| :  ||:||:|| |||| | |:|
Db       1017 SLEGRGWRLLAPITAYAQQTRGLFGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCING 1076

Qy       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
              | |||||||:|||| :||:||||:: : ||||| :| | :|| |||||: |||||||:
Db       1077 VCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWHAPQGARSLTPCTCGSSDLYLVTRH 1136

Qy       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200
              ||||| ||||| ||:||||||:| ||||||||:||| || ||:||||||:||||::||:
Db       1137 ADVIPVRRRGDSRGSLLSPRPISYLKGSSGGPLLCPSGHVVGIFRAAVCTRGVAKAVDFV 1196

Qy       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260
              |||:::   ||| |:|||:|||||| :|| :|||||||||||| ||||||||||||
Db       1197 PVESMETTMRSPVFTDNSSPPAVPQAFQVAHLHAPTGSGKSTKVPAAYAAQGYKVLVLNP 1256

Qy       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320
              ||||||||||:|||||::|||||||:|||||||||||||||||||:|||||||:||
Db       1257 SVAATLGFGAYMSKAHGVDPNIRTGVRTITTGAPITYSTYGKFLADGGCSGGAYDIIMCD 1316

Qy       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380
              |||: |||||||||||||||||| || |||||||||||| |||||||:||   ||||||
Db       1317 ECHSTDSTTILGIGTVLDQAETAGARLVVLATATPPGSVTVPHPNIEEIALSNTGEIPFY 1376

Qy       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440
              |:|||:  |||||||||||||||||||| |  :|:::|||||||||||||   |:||||
Db       1377 GKAIPIETIKGGRHLIFCHSKKKCDELAAKLSALGIHAVAYYRGLDVSVIPASGNVVVVA 1436

Qy       1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500
              |||||||:|||||||||||  ||| |||||||||||| |  :|||||||||||||| |||
Db       1437 TDALMTGFTGDFDSVIDCNTCVTQTVDFSLDPTFTIETTTMPQDAVSRSQRRGRTSRGRR 1496

Qy       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560
              ||||:|: ||| |||||| |||||||| |||||||:||:|||||| ||||||||||||
Db       1497 GIYRFVTPGERPSGMFDSSVLCECYDAGCAWYELTPAETSVRLRAYLNTPGLPVCQDHLE 1556

Qy       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
              |||:|||||||||||||||:|:|| || |||||||||:|||||| ||||||||||
Db       1557 FWESVFTGLTHIDAHFLSQTKQAGDNFPYLVAYQATVCARAQAPPPSWDQMWKCLTRLKP 1616

Qy       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680
              || |||||||||:| |||||||:|||| || ||||:|||||| ||||||:|||||
Db       1617 TLHGPTPLLYRLGAVQNEVTLTHPITKYIMACMSADLEVVTSTWVLVGGVLAALAAYCLT 1676

Qy       1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740
              || | |:||: :: :  | ||:||||: |||||||||    ||:| ::||  | |  |||
Db       1677 TGSVVIVGRIILSGKPAVVPDREVLYQEFDEMEECASHLPYIEQGMQLAEQFKQKALGLL 1736

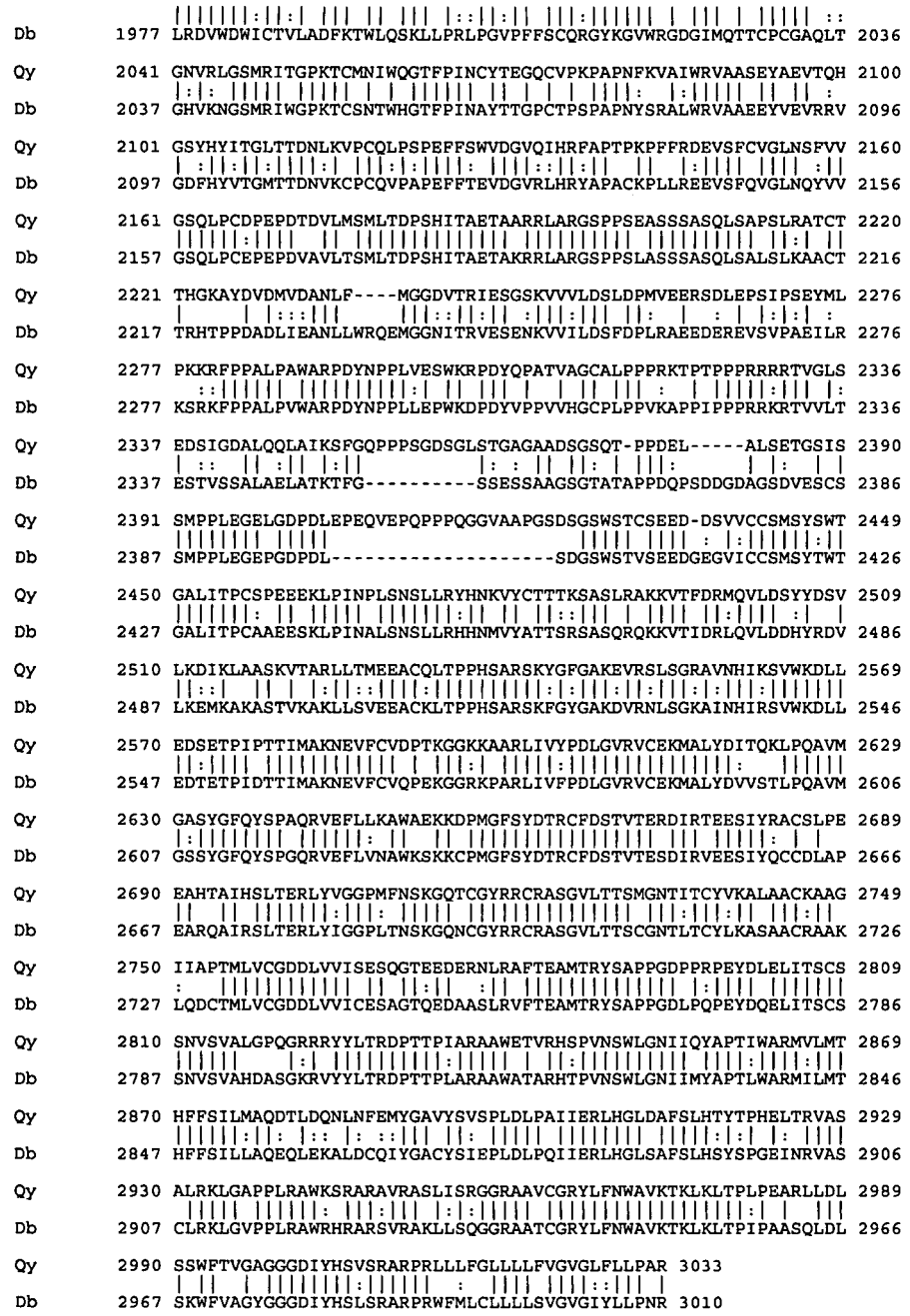
Qy       1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
              | |:|||:   | |:: |  :|  |||  |||||||||||||||||||:||:|||:|:
Db       1737 QTATKQAEAAAPVVESKWRTLEAFWANDMWNFISGIQYLAGLSTLPGNPAIASLMAFTAS 1796

Qy       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860
              :||||:| :|:| ||||||:|:|:|||  |: || :|: |||||||||||||::||||
Db       1797 ITSPLTTQSTLLFNILGGWVAAQLAPPGAASAFVGAGIAGAAVGSIGLGKVLVDMVAGYG 1856

Qy       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920
              ||::|||||||:|||| || ||:|||| ||||||||||:|||||||| |||||||||
Db       1857 AGVAGALVAFKVMSGEMPSTEDLVNLLPAILSPGALVVGVVCAAILRRHVDPGEGAVQWM 1916

Qy       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980
              |||||||||||||:||||| |||||: ||||:|  |||| ||||| || |||  || |||
Db       1917 NRLIAFASRGNHVSPTHYVPESDAAARVTQILSGLTITQLLRRLHQWINEDCSTPCSGSW 1976

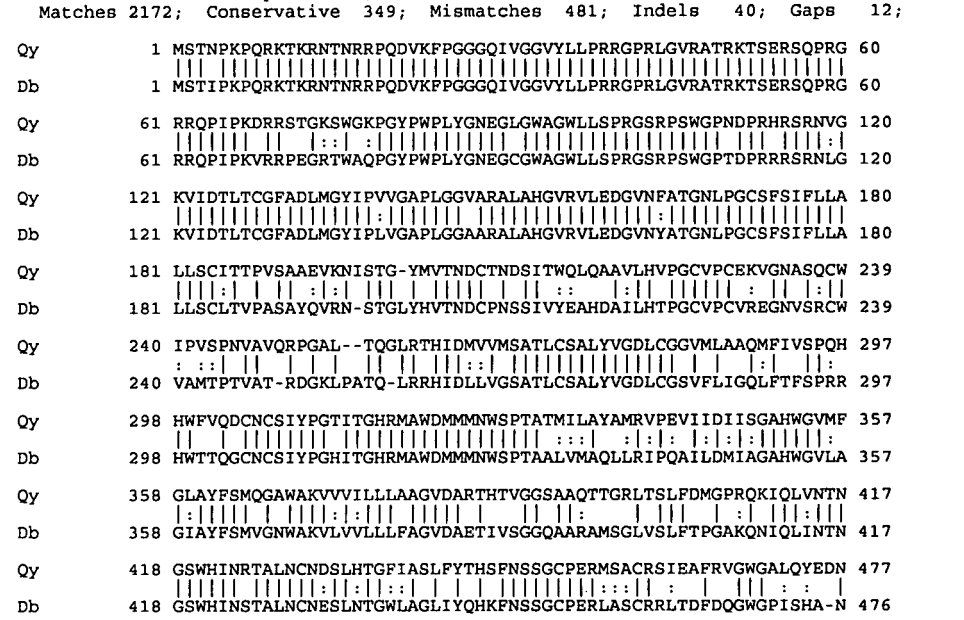
Qy       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040
```

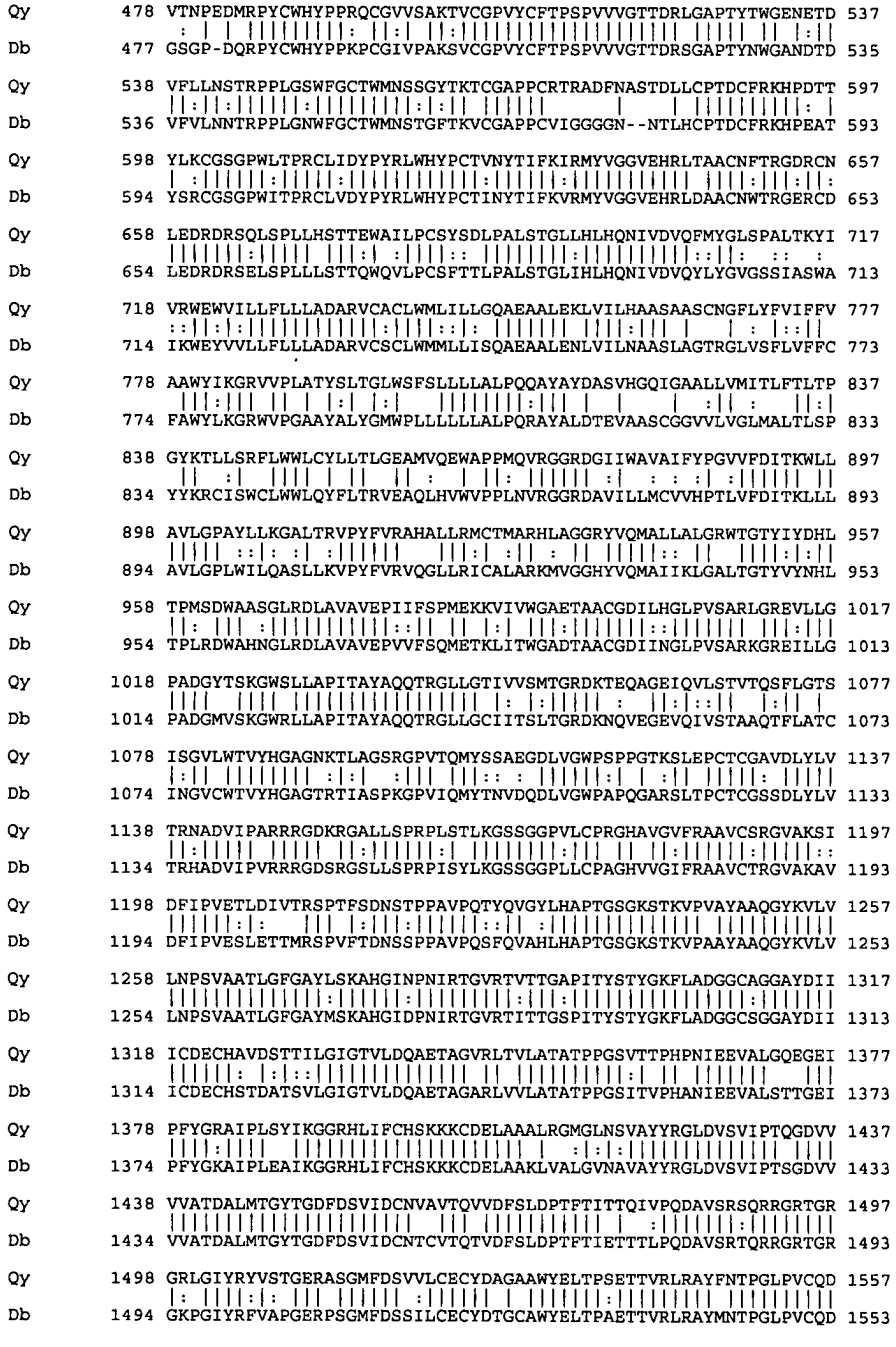


RESULT 8

S40770
genome polyprotein - hepatitis C virus
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 09-Jul-2004
C;Accession: S40770; PC1285
R;Okamoto, H.
submitted to the EMBL Data Library, March 1992
A;Reference number: S40770
A;Accession: S40770
A;Molecule type: genomic RNA
A;Residues: 1-3011 <OKA>
A;Cross-references: UNIPROT:Q03463; EMBL:D10749; NID:g221586; PIDN:BAA01582.1; PID:g2215
R;Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda,
Jpn. J. Exp. Med. 60, 167-177, 1990
A;Title: The 5'-terminal sequence of the hepatitis C virus genome.
A;Reference number: PC1284; MUID:91013116; PMID:2170712
A;Accession: PC1285
A;Molecule type: genomic RNA
A;Residues: 1-513 <OK2>
A;Cross-references: GB:D00831; NID:g221511; PIDN:BAA00705.1; PID:g221512
A;Experimental source: isolate HC-J1
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 73.8%; Score 11895; DB 1; Length 3011;
Best Local Similarity 71.4%; Pred. No. 0;
Matches 2172; Conservative 349; Mismatches 481; Indels 40; Gaps 12;




Qy 478 VTNPEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETD 537
Db 477 GSGP-DQRPYCWHYPPKPCGIVPAKSVCGPVYCFTPSPVVVGTTDRSGAPTYNWGANDTD 535
Qy 538 VFLLNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTT 597
Db 536 VFVLNNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVIGGGGN--NTLHCPTDCFRKHPEAT 593
Qy 598 YLKCGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCN 657
Db 594 YSRCGSGPWITPRCLVDYPYRLWHYPCTINYTIFKVRMYVGGVEHRLDAACNWTRGERCD 653
Qy 658 LEDRDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYI 717
Db 654 LEDRDRSELSPLLLSTTQWQVLPCSFTTLPALSTGLIHLHQNIVDVQYLYGVGSSIASWA 713
Qy 718 VRWEWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFV 777
Db 714 IKWEYVVLLFLLLADARVCSCLWMMLLISQAEAALENLVILNAASLAGTRGLVSFLVFFC 773
Qy 778 AAWYIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTP 837
Db 774 FAWYLKGRWVPGAAYALYGMWPLLLLLLALPQRAYALDTEVAASCGGVVLVGLMALTLSP 833
Qy 838 GYKTLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLL 897
Db 834 YYKRCISWCLWWLQYFLTRVEAQLHVWVPPLNVRGGRDAVILLMCVVHPTLVFDITKLLL 893
Qy 898 AVLGPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHL 957
Db 894 AVLGPLWILQASLLKVPYFVRVQGLLRICALARKMVGGHYVQMAIIKLGALTGTYVYNHL 953
Qy 958 TPMSDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLG 1017
Db 954 TPLRDWAHNGLRDLAVAVEPVVFSQMETKLITWGADTAACGDIINGLPVSARKGREILLG 1013
Qy 1018 PADGYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTS 1077
Db 1014 PADGMVSKGWRLLAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATC 1073
Qy 1078 ISGVLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLV 1137
Db 1074 INGVCWTVYHGAGTRTIASPKGPVIQMYTNVDQDLVGWPAPQGARSLTPCTCGSSDLYLV 1133
Qy 1138 TRNADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSI 1197
Db 1134 TRHADVIPVRRRGDSRGSLLSPRPISYLKGSSGGPLLCPAGHVVGIFRAAVCTRGVAKAV 1193
Qy 1198 DFIPVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLV 1257
Db 1194 DFIPVESLETTMRSPVFTDNSSPPAVPQSFQVAHLHAPTGSGKSTKVPAAYAAQGYKVLV 1253
Qy 1258 LNPSVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDII 1317
Db 1254 LNPSVAATLGFGAYMSKAHGIDPNIRTGVRTITTGSPITYSTYGKFLADGGCSGGAYDII 1313
Qy 1318 ICDECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEI 1377
Db 1314 ICDECHSTDATSVLGIGTVLDQAETAGARLVVLATATPPGSITVPHANIEEVALSTTGEI 1373
Qy 1378 PFYGRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVV 1437
Db 1374 PFYGKAIPLEAIKGGRHLIFCHSKKKCDELAAKLVALGVNAVAYYRGLDVSVIPTSGDVV 1433
Qy 1438 VVATDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGR 1497
Db 1434 VVATDALMTGYTGDFDSVIDCNTCVTQTVDFSLDPTFTIETTTLPQDAVSRTQRRGRTGR 1493
Qy 1498 GRLGIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQD 1557
Db 1494 GKPGIYRFVAPGERPSGMFDSSILCECYDTGCAWYELTPAETTVRLRAYMNTPGLPVCQD 1553
Qy 1558 HLEFWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTR 1617
Db 1554 HLEFWEGVFTGLTHIDAHFLSQTKQGGENFPYLVAYQATVCARAQAPPPSWDQMWKCLIR 1613
Qy 1618 LKPTLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAY 1677
Db 1614 LKPTLHGPTPLLYRLGAVQGEVTLTHPVTKYIMTCMSADLEVVTSTWVLVGGVLAALAAY 1673
Qy 1678 CLATGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQ 1737
Db 1674 CLSTGCVVIVGRIVLSGRPAIIPDREVLYREFDEMEECSQHLPYIEQGMMLAEQFKQKAL 1733
Qy 1738 GLLQQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAF 1797
Db 1734 GLLQTASRQAEVIAPTVQTNWQKLEAFWAKHMWNFISGIQYLAGLSTLPGNPAIASLMAF 1793
Qy 1798 SAALTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILA 1857
Db 1794 TAAVTSPLTTSQTLLFNILGGWVAAQLAAPGAATAFVGSGLAGAAVGSVGLGRVLVDILA 1853
Qy 1858 GYGAGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAV 1917
Db 1854 GYGAGVAGALVAFKIMSGELPSTEDLVNLLPAILSPGALVVGVVCAAILRRHVGPGEGAV 1913
Qy 1918 QWMNRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCG 1977
Db 1914 QWMNRLIAFASRGNHVSPTHYVPESDAAARVTAILSSLTVTQLLRRLHQWLSSESTTPCS 1973
Qy 1978 GSWLRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGA 2037
Db 1974 GSWLRDIWDWICEVLSDFKTWLKTKLMPHLPGIPFVSCQHGYKGVWRGDGIMHTRCHCGA 2033
Qy 2038 NISGNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEV 2097
Db 2034 EITGHVKNGTMRIVGPKTCRNMWSGTFPINAYTTGPCTPLPAPNYTFALWRVSAEEYVEI 2093
Qy 2098 TQHGSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNS 2157
Db 2094 RRVGDFHYVTGMTTDNLKCPCQVPSPEFFTELDGVRLHRFAPPCKPLLREEVSFRVGLHD 2153
Qy 2158 FVVGSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRA 2217
Db 2154 YPVGSQLPCEPEPDVAVLTSMLTDPSHITAAAAGRRLARGSPPSEASSSASQLSAPSLKA 2213
Qy 2218 TCTTHGKAYDVDMVDANLF----MGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSE 2273
Db 2214 TCTINHDSPDAELIEANLLWRQEMGGNITRVESENKVVILDSFDPLVAEEDEREISVPAE 2273
Qy 2274 YMLPKKRFPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTV 2333
Db 2274 ILRKSRRFTQALPIWARPDYNPPLIETWKKPNYEPPVVHGCPLPPPQSPPVPPPRKKRTV 2333
Qy 2334 GLSEDSIGDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTPPDELALSETGSISSMP 2393
Db 2334 VLTESTLSTALAELAAKSFGSSSTSGITGDNTTTSSEPAPSGCSPD----SDAESYSSMP 2389
Qy 2394 PLEGELGDPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEE--DDSVVCCSMSYSWTGA 2451
Db 2390 PLEGEPGDPDL-------------------SDGSWSTVSSEAGTEDVVCCSMSYTWTGA 2429
Qy 2452 LITPCSPEEEKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLK 2511
Db 2430 LITPCAAEEQKLPINALSNSLLRHHNLVYSTTSRSACQRQKKVTFDRLQVLDSHYQDVLK 2489
Qy 2512 DIKLAASKVTARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLED 2571
Db 2490 EVKAAASKVKANLLSVEEACSLTPPHSAKSKFGYGAKDVRCHARKAVNHINSVWKDLLED 2549
Qy 2572 SETPIPTTIMAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGA 2631
Db 2550 SVTPIQTTIMAKNEVFCVQPEKGGRKPARLIVFPDLGVRVCEKMALYDVVSKLPPAVMGS 2609
Qy 2632 SYGFQYSPAQRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEA 2691

```
Db       2610 SYGFQYSPGQRVEFLVQAWKSKRTPMGFSYDTRCFDSTVTESDIRTEEAIYQCCDLDPQA 2669

Qy       2692 HTAIHSLTERLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGII 2751

Db       2670 RVAIRSLTERLYVGGPLTNSRGENCGYRRCRASGVLTTSCGNTLTCYIKARAACRAAGLQ 2729

Qy       2752 APTMLVCGDDLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSN 2811

Db       2730 DCTMLVCGDDLVVICESAGVQEDAASLRAFTEAMTRYSAPPGDPPQPEYDLELITSCSSN 2789

Qy       2812 VSVALGPQGRRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHF 2871

Db       2790 VSVAHDGTGKRVYYLTRDPTTPLARAAWETARHTPVNSWLGNIIMFAPTLWARMILMTHF 2849

Qy       2872 FSILMAQDTLDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASAL 2931

Db       2850 FSVLIARDQLEQALDCEIYGACYSIEPLDLPPIIQRLHGLSAFSLHSYSPGEINRVAACL 2909

Qy       2932 RKLGAPPLRAWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSS 2991

Db       2910 RKLGVPPLRAWRHRARSVRARLLSRGGRAAICGKYLFNWAVRTKLKLTPIAAAGRLDLSG 2969

Qy       2992 WFTVGAGGGDIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033

Db       2970 WFTAGYSGGDIYHSVSHARPRWFWFCLLLLAAGVGIYLLPNR 3011


RESULT 9
S18030
genome polyprotein - hepatitis C virus (isolate JK1)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
A;Variety: isolate JK1
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 09-Jul-2004
C;Accession: S18030; S33570; A48332; S18029
R;Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
submitted to the EMBL Data Library, September 1991
A;Description: A whole genome of hepatitis C virus cDNA was isolated from a single patie
A;Reference number: S18028
A;Accession: S18030
A;Molecule type: genomic RNA
A;Residues: 1-3010 <HON>
A;Cross-references: UNIPROT:Q68949; EMBL:X61596; NID:g59478; PIDN:CAA43793.1; PID:g59479
A;Experimental source: isolate JK1 from an individual
R;Honda, M.; Kaneko, S.; Unoura, M.; Kobayashi, K.; Murakami, S.
Arch. Virol. 128, 163-169, 1993
A;Title: Sequence analysis of putative structural regions of hepatitis C virus isolated
A;Reference number: A48332; MUID:93119270; PMID:8380322
A;Accession: S33570
A;Molecule type: genomic RNA
A;Residues: 1-547,'T',549-621,'V',623-624,'S',626-652,'DL',655-761,'T',763-782 <HOW>
A;Cross-references: EMBL:X61591
A;Note: this sequence is inconsistent with the nucleotide translation
A;Note: the authors translated the codon AGG for residue 43 as Pro, TGG for residue 320
 as Trp, and TTC for residue 771 as Ser
A;Note: sequence extracted from NCBI backbone (NCBIN:121747, NCBIP:121748)
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate (As
```

```
  Query Match             73.5%;  Score 11843.5;  DB 1;  Length 3010;
  Best Local Similarity   70.9%;  Pred. No. 0;
  Matches 2157;  Conservative  346;  Mismatches  496;  Indels   45;  Gaps    8;

Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120

Db         61 RRQPIPKARQPEGRAWAQPGYPWPLYGNEGLGWAGWLLSPYGSRPRWGPTDPRRRSRNLG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180

Db        121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240

Db        181 LLSCLTVPVSTYEVRNVSGVYHVTNDCSNSSIVYEAADMIMHTPGCVPCVREGNSSRCWV 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300

Db        241 ALTPTLAARNSSIPTTTIRRHVDLLVGAAALCSAMYVGDLCGSVFLVSQLFTFSPRRYET 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360

Db        301 VQDCNCSLYPGHVSGHRMAWDMMMNWSPTTALVVSQLLRIPQAVVDMVVGAHWGVLAGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420

Db        361 YYSMVGNWAKVLIVMLLFAGVDGTTYVSVGHASQTTRRVASFFSPGSAQKIQLVNTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480

Db        421 HINRTALNCNESINTGFFAALFYVKKFNSSGCSERMASCRPIDRFAQGWGPITHAE--SR 478

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540

Db        479 SSDQRPYCWHYAPQPCGIVPALQVCGPVYCFTPSPVVVGTTDRFGVPTYNWGDNETDVLL 538

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600

Db        539 LNNTRPPQGNWFGCTWMNSTGFTKTCGGPPCNIGGAGN--NTLTCPTDCFRKHPEATYTK 596

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660

Db        597 CGSGPWLTPRCMVDYPYRLWHYPCTFNFTIFKIRMYVGGVEHRLNAACNWTRGERCNIED 656

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720

Db        657 RDRSELSPLLLSTTEWQILPCSFTTLPALSTGLIHLHQNIVDVQYLYGVGSAVVSIVIKW 716

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780

Db        717 EYVLLLFLLLADARVCACLWMMLLIAQAEAALENLVVLNAASVAGAHGILSFLVFFCAAW 776

Qy        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840

Db        777 YIKGRLVPGAAYAFYGVWPLLLLLLALPPRAYAMDREMAASCGGAVFVGLALLTLSPHYK 836

Qy        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900

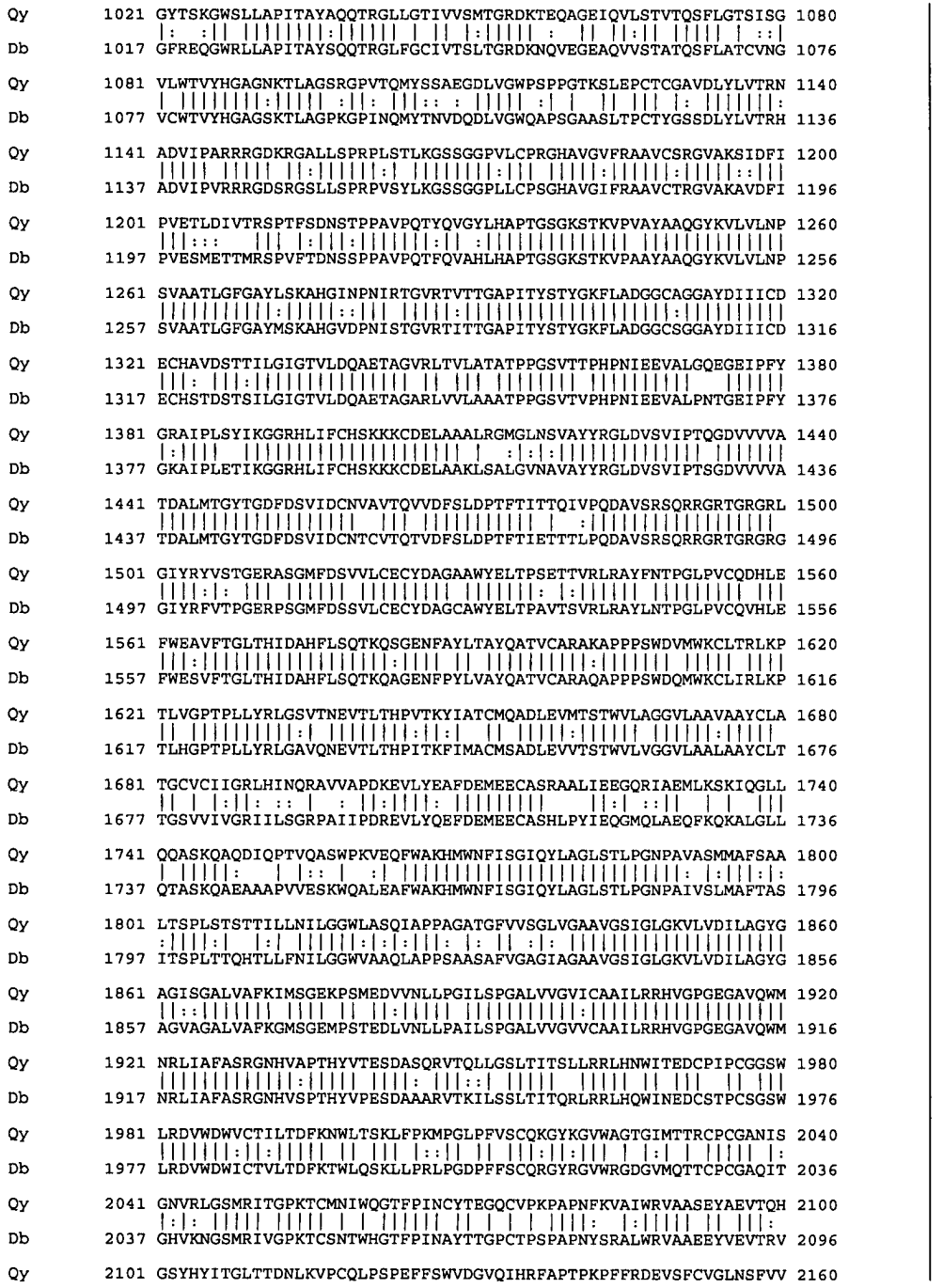
Db        837 AFLARLIWWLQYFITRVEACLQVWVPPLNVRGGRDAIILLTCAVHSELIFDITKILLAIF 896

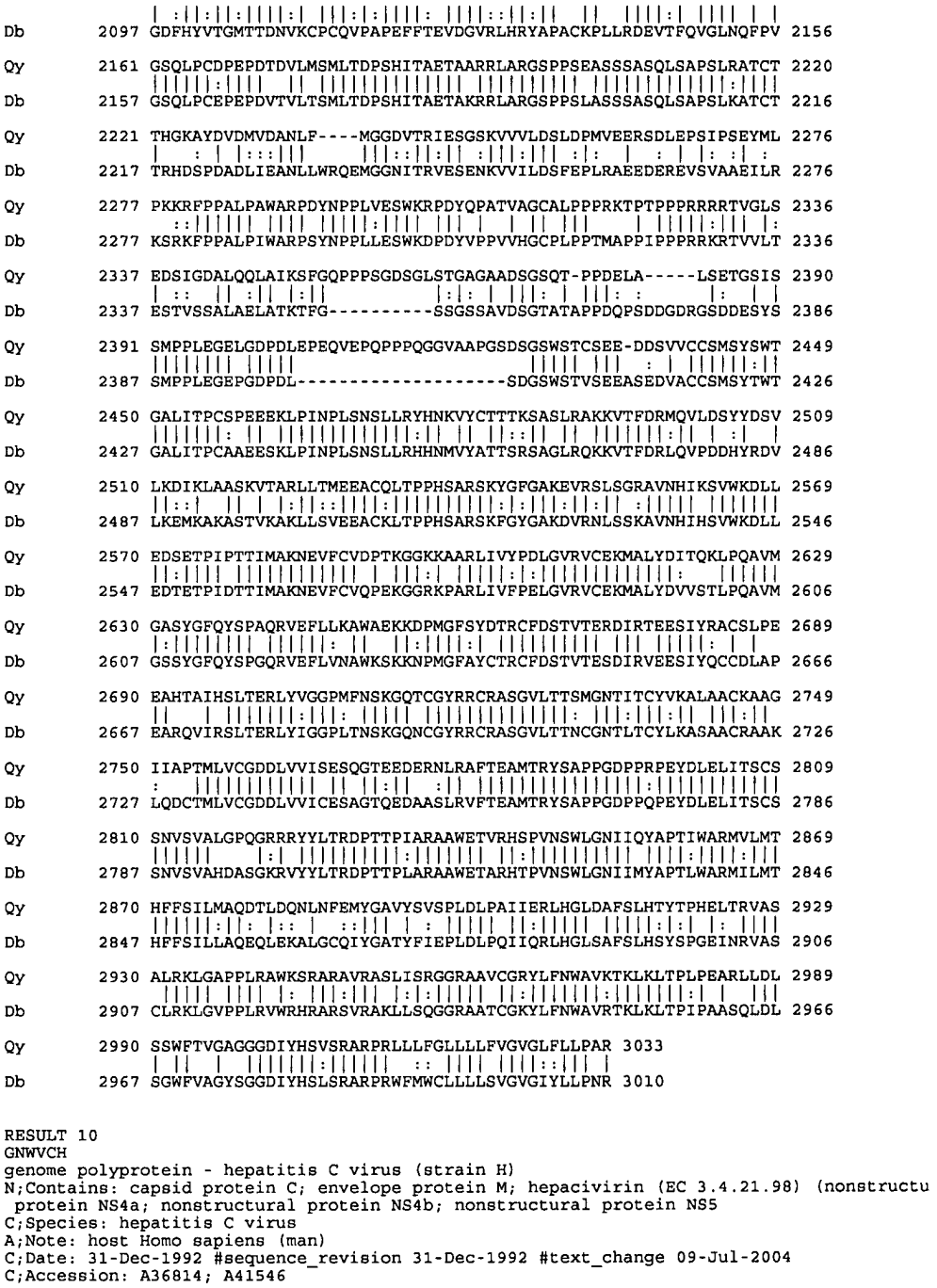
Qy        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960

Db        897 GPLMVLQAGLTRVPYFVRAQGLLRACMLVRKVAGGHYVQMALMKLAALTGTYVYDHLTPL 956

Qy        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020

Db        957 RDWAHSGLRDLVVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGREILLGPAD 1016
```





RESULT 10
GNWVCH
genome polyprotein - hepatitis C virus (strain H)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
A;Note: host Homo sapiens (man)
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 09-Jul-2004
C;Accession: A36814; A41546

```
R;Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
submitted to GenBank, July 1992
A;Description: Genomic structure of the human prototype strain H of hepatitis C virus: c
A;Reference number: A36814
A;Accession: A36814
A;Molecule type: genomic RNA
A;Residues: 1-3011 <INC>
A;Cross-references: UNIPROT:P27958; GB:M67463; NID:g329737; PIDN:AAA45534.1; PID:g329738
R;Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991
A;Title: Genomic structure of the human prototype strain H of hepatitis C virus: compari
A;Reference number: A41546; MUID:92052256; PMID:1658800
A;Contents: annotation
A;Note: neither amino acid nor nucleotide sequence is given
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
F;1-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,23

  Query Match           73.2%;  Score 11796;  DB 1;  Length 3011;
  Best Local Similarity  70.7%;  Pred. No. 0;
  Matches 2149;  Conservative  364;  Mismatches  491;  Indels   36;  Gaps    8;

Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              |||||||  ||  |::| :|||||||||||||| ||||||||||||||||||| ||| ||||:|
Db         61 RRQPIPKARRPEGRTWAQPGYPWPLYGNEGCGWAGWLLSPRGSRPSWGPTDPRRRSRNLG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
              |||||||||||||||||:||||||| |||||||||||||:|||||||||||||||||||||
Db        121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
              ||||:| | || :|:| |  | ||||| | |: ::   |:|| |||||| : ||||:||:
Db        181 LLSCLTVPASAYQVRNSSGLYHVTNDCPNSSVVYEAADAILHTPGCVPCVREGNASRCWV 240

Qy        241 PVSPNVAVQRPGAL-TQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHW 299
              |:| ||  | | | |  || |||::| |||||||||||||||| | |  |:|  ||:|||
Db        241 AVTPTVAT-RDGKLPTTQLRRHIDLLVGSATLCSALYVGDLCGSVFLVGQLFTFSPRHHW 299

Qy        300 FVQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGL 359
              |||||||||| |||||||:|||||||| :::|  :|:|: |:|:|:|||||: |:
Db        300 TTQDCNCSIYPGHITGHRMAWNMMMNWSPTAALVVAQLLRIPQAIMDMIAGAHWGVLAGI 359

Qy        360 AYFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGS 419
              |||| | ||||:|:||| ||||| ||  ||:| :||  |  |   | :| |||:|||||
Db        360 KYFSMVGNWAKVLVVLLLFAGVDAETHVTGGNAGRTTAGLVGLLTPGAKQNIQLINTNGS 419

Qy        420 WHINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVT 479
              |||| ||||||:||:||::| ||| | |||||||||:::|| :  |  ||| : | :
Db        420 WHINSTALNCNESLNTGWLAGLFYQHKFNSSGCPERLASCRRLTDFAQGWGPISYANG-- 477

Qy        480 NPEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVF 539
              :  | |||||||||| ||:| ||:|||||||||||||||||| |||||:|| |:||||
Db        478 SGLDERPYCWHYPPRPCGIVPAKSVCGPVYCFTPSPVVVGTTDRSGAPTYSWGANDTDVF 537

Qy        540 LLNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYL 599
              :||:||||||:|||||||:|:|| ||||||      |    ||||||||||:|: ||
Db        538 VLNNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVIGGVGN--NTLLCPTDCFRKYPEATYS 595

Qy        600 KCGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLE 659
              :||||| :||||::|||||||||||||:||||||:|||||||||| |||||:|||:||:||
Db        596 RCGSGPRITPRCMVDYPYRLWHYPCTINYTIFKVRMYVGGVEHRLEAACNWTRGERCDLE 655

Qy        660 DRDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVR 719
              |||||:||||| |||:|| :||||:: |||||||||:||||||||::||:  ::  : ::
Db        656 DRDRSELSPLLLSTTQWQVLPCSFTTLPALSTGLIHLHQNIVDVQYLYGVGSSIASWAIK 715

Qy        720 WEWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAA 779
              ||:|:|||||||||||||:||||::|: |||||||  |||||||| |  :| : |::||  |
Db        716 WEYVVLLFLLLADARVCSCLWMMLLISQAEAALENLVILNAASLAGTHGLVSFLVFFCFA 775

Qy        780 WYIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGY 839
              ||:||| || | |:| |:|   |||||||||:||| |  |     |  :|| :   ||:| |
Db        776 WYLKGRWVPGAVYALYGMWPLLLLLLALPQRAYALDTEVAASCGGVVLVGLMALTLSPYY 835

Qy        840 KTLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAV 899
              |  :|  :||| | ||  || :  | ||: |||||| :|    : :| :|||||| |||:
Db        836 KRYISWCMWWLQYFLTRVEAQLHVWVPPLNVRGGRDAVILLTCVVHPALVFDITKLLLAI 895

Qy        900 LGPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTP 959
              || ::|: :| :|||||   |||:| :|| :||| |||||:: ||  ||| :|:|| |
Db        896 FGPLWILQASLLKVPYFVRVQGLLRICALARKIAGGHYVQMAIIKLGALTGTCVYNHLAP 955

Qy        960 MSDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPA 1019
              : ||| :||||||||||::|| || |:| |||:|||||||::|||||| |:|:|||||
Db        956 LRDWAHNGLRDLAVAVEPVVFSRMETKLITWGADTAACGDIINGLPVSARRGQEILLGPA 1015

Qy       1020 DGYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSIS 1079
              ||  |||| ||||||||||||||||| |: |:||||| :  ||:|::|| ||:|| | |:
Db       1016 DGMVSKGWRLLAPITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTATQTFLATCIN 1075

Qy       1080 GVLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTR 1139
              || |||||||  :|:|  :||| | |:: : |||||:|  |::|| |||||: |||||||
Db       1076 GVCWTVYHGAGTRTIASPKGPVIQTYTNVDQDLVGWPAPQGSRSLTPCTCGSSDLYLVTR 1135

Qy       1140 NADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDF 1199
              :||||| ||||| ||:|||||||| |||||||:||| |||||:|||||:|||||::||
Db       1136 HADVIPVRRRGDSRGSLLSPRPISYLKGSSGGPLLCPTGHAVGLFRAAVCTRGVAKAVDF 1195

Qy       1200 IPVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLN 1259
              |||| |:   ||| |:|||:||||||::|| :|||||||||||||| ||||:||||||||
Db       1196 IPVENLETTMRSPVFTDNSSPPAVPQSFQVAHLHAPTGSGKSTKVPAAYAAKGYKVLVLN 1255

Qy       1260 PSVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIIC 1319
              |||||||||||:|||||::|||||||:|||:|||||||||||||| ||:|||||||||||
Db       1256 PSVAATLGFGAYMSKAHGVDPNIRTGVRTITTGSPITYSTYGKFLADAGCSGGAYDIIIC 1315

Qy       1320 DECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPF 1379
              ||||: |:|:| ||||||||||||| || |||||||||||   |||||||||   |||||
Db       1316 DECHSTDATSISGIGTVLDQAETAGARLVVLATATPPGSVTVSHPNIEEVALSTTGEIPF 1375

Qy       1380 YGRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVV 1439
              ||:|||| |||||||||||||||||||| |  :|:|:|||||||||||||| ||||||
Db       1376 YGKAIPLEVIKGGRHLIFCHSKKKCDELAAKLVALGINAVAYYRGLDVSVIPTSGDVVVV 1435

Qy       1440 ATDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGR 1499
              :|||||:|||||||||| |||||||||| |  :|||||||:|||||||:
Db       1436 STDALMTGFTGDFDSVIDCNTCVTQTVDFSLDPTFTIETTTLPQDAVSRTQRRGRTGRGK 1495

Qy       1500 LGIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHL 1559
              ||||:|: ||| |||||| |||||||| |||||||:|||||||| ||||||||||||
Db       1496 PGIYRFVAPGERPSGMFDSSVLCECYDAGCAWYELTPAETTVRLRAYMNTPGLPVCQDHL 1555

Qy       1560 EFWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLK 1619
              ||| |||||||||||||||||||| || |||||||||:|||||| | ||| |||
Db       1556 GFWEGVFTGLTHIDAHFLSQTKQSGENFPYLVAYQATVCARAQAPPPSWDQMRKCLIRLK 1615

Qy       1620 PTLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCL 1679
              ||| ||||||||||:| |||||||:|||| ||| |||||:|||||| |||||||||||
Db       1616 PTLHGPTPLLYRLGAVQNEVTLTHPITKYIMTCMSADLEVVTSTWVLVGGVLAALAAYCL 1675
```

```
Qy       1680 ATGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGL 1739
             :|||| |:||: :: :  : ||:||||: |||||||:     ||:|  :||  | |  ||
Db       1676 STGCVVIVGRIVLSGKPAIIPDREVLYQEFDEMEECSQHLPYIEQGMMLAEQFKQKALGL 1735

Qy       1740 LQQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSA 1799
             || ||: |: | | || :| |:| |||||||||||||||||||||||||||:||:|||:|
Db       1736 LQTASRHAEVITPAVQTNWQKLEVFWAKHMWNFISGIQYLAGLSTLPGNPAIASLMAFTA 1795

Qy       1800 ALTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGY 1859
             |:||||:|  |:| ||||||:|:|:| |  || || :|| |||: |:||||||||||||
Db       1796 AVTSPLTTGQTLLFNILGGWVAAQLAAPGAATAFVGAGLAGAALDSVGLGKVLVDILAGY 1855

Qy       1860 GAGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQW 1919
             |||::|||||||||||| || ||:||||| ||||||| |||: |:|||| ||||||||||
Db       1856 GAGVAGALVAFKIMSGEVPSTEDLVNLLPAILSPGALAVGVVFASILRRRVGPGEGAVQW 1915

Qy       1920 MNRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGS 1979
             ||||||||||||||:||||| |||||: ||| :| |||:| |||||| ||: :|  || ||
Db       1916 MNRLIAFASRGNHVSPTHYVPESDAAARVTAILSSLTVTQLLRRLHQWISSECTTPCSGS 1975

Qy       1980 WLRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANI 2039
             ||||:|||:| :|:||| || :|| |::||:||||||:||:||| | ||| ||| ||| |
Db       1976 WLRDIWDWICEVLSDFKTWLKAKLMPQLPGIPFVSCQRGYRGVWRGDGIMHTRCHCGAEI 2035

Qy       2040 SGNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQ 2099
             :|:|: |:||| ||:|| |:| ||| || || | | ||||:| |:|||:| || |: :
Db       2036 TGHVKNGTMRIVGPRTCKNMWSGTFFINAYTTGPCTPLPAPNYKFALWRVSAEEYVEIRR 2095

Qy       2100 HGSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFV 2159
               | :||::|:|||||| |||:|||||: :|||::|||||  ||  |:|||| |||: :
Db       2096 VGDFHYVSGMTTDNLKCPCQIPSPEFFTELDGVRLHRFAPPCKPLLREEVSFRVGLHEYP 2155

Qy       2160 VGSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATC 2219
             |||||||:|||| || |||||||||||| | |||||||||| ||||||||||||:|||
Db       2156 VGSQLPCEPEPDVAVLTSMLTDPSHITAEAAGRRLARGSPPSMASSSASQLSAPSLKATC 2215

Qy       2220 TTHGKAYDVDMVDANLF----MGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYM 2275
             | :  : | ::::|||     |||::||:|| :|||:||| ||:| |  : | |:|:| :
Db       2216 TANHDSPDAELIEANLLWRQEMGGNITRVESENKVVILDSFDPLVAEEDEREVSVPAEIL 2275

Qy       2276 LPKKRFPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGL 2335
                :|| |||| |||||||| |||:||:|||:|  | || |||||  | ||||::||| |
Db       2276 RKSRRFAPALPVWARPDYNPLLVETWKKPDYEPPVVHGCPLPPPRSPPVPPPRKKRTVVL 2335

Qy       2336 SEDSIGDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTPPDELALSETGSISSMPPL 2395
             :| ::  || :|| |||     || :| :|   :  : | |||     |:  | ||||||
Db       2336 TESTLPTALAELATKSFGSSSTSGITGDNTTTSSEPAPSGCPPD----SDVESYSSMPPL 2391

Qy       2396 EGELGDPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSE--EDDSVVCCSMSYSWTGALI 2453
             ||| |||||                    |||||  |   : : |||||||||||||:
Db       2392 EGEPGDPDL--------------------SDGSWSTVSSGADTEDVVCCSMSYSWTGALV 2431

Qy       2454 TPCSPEEEKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDI 2513
             |||: ||:||||| |||||||:|| || ||::||  | |||||||:|||||:|  |||::
Db       2432 TPCAAEEQKLPINALSNSLLRHHNLVYSTTSRSACQRKKKVTFDRLQVLDSHYQDVLKEV 2491

Qy       2514 KLAASKVTARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSE 2573
             | ||||| | ||::|||| | |||||:||:|:|||:||  : :|| || ||||||||||
Db       2492 KAAASKVKANLLSVEEACSLAPPHSAKSKFGYGAKDVRCHARKAVAHINSVWKDLLEDSV 2551

Qy       2574 TPIPTTIMAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASY 2633
             ||| |||||||||||| | |||:| |||||:|||||||||||||||:  ||| ||||:||
Db       2552 TPIDTTIMAKNEVFCVQPEKGGRKPARLIVFPDLGVRVCEKMALYDVVSKLPLAVMGSSY 2611

Qy       2634 GFQYSPAQRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHT 2693
             |||||| |||||::||  || ||| |||||||||||||| ||||||:||: | |  :|
Db       2612 GFQYSPGQRVEFLVQAWKSKKTPMGLSYDTRCFDSTVTESDIRTEEAIYQCCDLDPQARV 2671

Qy       2694 AIHSLTERLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAP 2753
             || |||||||||: ||:|: |||||||| ||||| |||:| |:|| |||:|||:
Db       2672 AIKSLTERLYVGGPLTNSRGENCGYRRCRASRVLTTSCGNTLTRYIKARAACRAAGLQDC 2731

Qy       2754 TMLVCGDDLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVS 2813
             |||||||||||| || | :||  :||||||||||||||||||:|||||||||||||||
Db       2732 TMLVCGDDLVVICESAGVQEDAASLRAFTEAMTRYSAPPGDPPQPEYDLELITSCSSNVS 2791

Qy       2814 VALGPQGRRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFS 2873
             ||    |:| |||||||||:||||||| ||:|||||||||| :|||:||||:|||||||
Db       2792 VAHDGAGKRVYYLTRDPTTPLARAAWETARHTPVNSWLGNIIMFAPTLWARMILMTHFFS 2851

Qy       2874 ILMAQDTLDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRK 2933
             :|:|:| |:| || |:||| ||: ||||| ||:||||| |||||:|:| |: |||: |||
Db       2852 VLIARDQLEQALNCEIYGACYSIEPLDLPPIIQRLHGLSAFSLHSYSPGEINRVAACLRK 2911

Qy       2934 LGAPPLRAWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWF 2993
             || ||||||: || :||| |::|||:||:||:|||||||:|||||||:   |  |||| ||
Db       2912 LGVPPLRAWRHRAWSVRARLLARGGKAAICGKYLFNWAVRTKLKLTPITAAGRLDLSGWF 2971

Qy       2994 TVGAGGGDIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
             | |  |||||||| ||||   | ||||  |||::||| |
Db       2972 TAGYSGGDIYHSVSHARPRWFWFCLLLLAAGVGIYLLPNR 3011
```

RESULT 11
JC5620
genome polyprotein - hepatitis C virus (isolate EUH1480)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 09-Jul-2004
C;Accession: JC5620
R;Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.
Biochem. Biophys. Res. Commun. 236, 44-49, 1997
A;Title: The complete coding sequence of hepatitis C virus genotype 5a, the predominant
A;Reference number: JC5620; MUID:97366593; PMID:9223423
A;Accession: JC5620
A;Molecule type: mRNA
A;Residues: 1-3014 <CHA>
A;Cross-references: UNIPROT:O39928; GB:Y13184
A;Experimental source: genotype 5a, which predominates in South Africa
A;Note: the translation of the nucleotide sequence is not complete in this paper
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;384-408/Region: hypervariable #status predicted
F;390-730/Product: nonstructural protein NS1 #status predicted <NS1>
F;731-1007/Product: nonstructural protein NS2 #status predicted <NS2>
F;1008-1616/Product: hepacivirin #status predicted <NS3>
F;1231-1238/Region: nucleotide-binding motif A (P-loop)
F;1313-1318/Region: nucleotide-binding motif B
F;1317-1320/Region: DEXH motif
F;1617-1863/Product: nonstructural protein NS4a #status predicted <N4A>
F;1864-2014/Product: nonstructural protein NS4b #status predicted <N4B>
F;2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>
F;2210-2249/Region: interferon sensitivity determining #status predicted

 Query Match 71.3%; Score 11500.5; DB 1; Length 3014;
 Best Local Similarity 69.0%; Pred. No. 0;
 Matches 2099; Conservative 363; Mismatches 544; Indels 35; Gaps 9;

```
Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              |||||||||||||||||||||||||||||||||||||||||||||:|||||||| ||||||||
Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPKLGVRATRKNSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
             ||||||| || ||:|||:|||||||| |||||||||||||| |||:||||||| :| |:|
Db         61 RRQPIPKARRPTGRSWGQPGYPWPLYANEGLGWAGWLLSPRSSRPNWGPNDPRRKSPNLG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
             :|| ||||||  ||||||:|| |:|||:|||||||:|||||:|:|||||||| ||||:||
Db        121 RVIHTLTCGFPHLMGYIPLVGGPVGGVSRALAHGVKVLEDGINYATGNLPGCPFSIFVLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
             || |:| | ||   :| |  | ||||| | || ::    :|| ||||||    | |:||:
```

```
Db        181 LLWCLTVPASAVPYRNASGVYHVTNDCPNSSIVYEADNLILHAPGCVPCVLEDNVSRCWV 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300

Db        241 QITPTLSAPSFGAVTALLRRAVDYLAGGAAFCSALYVGDACGALSLVGQMFTYKPRQHTT 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360

Db        301 VQDCNCSIYSGHITGHRMAWDMMMKWSPTTALLMAQLLRIPQVVIDIIAGGHWGVLLAAA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420

Db        361 YFASTANWAKVILVLFLFAGVDGRTHTVGGTVGQGLKSLTSFFNPGPQRQLQFVNTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480

Db        421 HINSTALNCNDSLQTGFIAGLMYAHKFNSSGCPERMSSCRPLAAFDQGWGTISYA-TISG 479

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540

Db        480 PSDDKPYCWHYPPRPCGVVPARDVCGPVYCFTPSPVVVGTTDRRGCPTYNWGSNETDILL 539

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600

Db        540 LNNIRPPAGNWFGCTWMNSTGFVKNCGAPPCNLGPTGNNS--LKCPTDCFRKHPDATYTR 597

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660

Db        598 CGSGPWLTPRCLVHYPYRLWHYPCTVNYTIFKVRMFIGGLEHRLEAACNWTYGERCDLED 657

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720

Db        658 RDRAELSPLLHTTTQWAILPCSFTPTPALSTGLIHLHQNIVDTQYLYGLSSSIVSWAVKW 717

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780

Db        718 EYIMLVFLLLADARICTCLLILLLICQAEATCKNVIVLNAAAAAAGNHGFFWGLLVVCLAW 777

Qy        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840

Db        778 HVKGRLVPGATYLCLGVWPLLLVRLLRPHRALALDSSDGGTVGCLVLIVLTIFTLTPGYK 837

Qy        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900

Db        838 KKVVLVMWWLQYFIARVEAIIHVWVPPLQVKGGRDAVIMLTCLFHPALGFEITKILFGIL 897

Qy        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960

Db        898 GPLYLLQHSLTKVPYFLRARALLRLCLLAKHLVYGKYVQAALLHLGRLTGTYIYDHLAPM 957

Qy        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020

Db        958 KDWAASGLRELTVATEPIVFSAMETKVITWGADTAACGNILAVLPVSARRGREIFLGPAD 1017

Qy       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080

Db       1018 DIKTSGWRLLAPITAYAQQTRGVLGAIVLSLTGRDKNEAEGEVQFLSTATQTFLGICING 1077

Qy       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140

Db       1078 VMWTLFHGAGSKTLAGPKGPVVQMYTNVDKDLVGWPSPPGKGSLTRCTCGSADLYLVTRH 1137

Qy       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200

Db       1138 ADVIPARRRGDTRASLLSPRPISYLKGSSGGPIMCPSGHVVGVFRAAVCTRGVAKALEFV 1197

Qy       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260

Db       1198 PVENLETTMRSPVFTDNSTPPAVPHEFQVGHLHAPTGSGKSTKVPAAYAAQGYKVLVLNP 1257

Qy       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320

Db       1258 SVAATFGFGAYMSRAYGVDPNIRTGVRTVTTGAGITYSTYGKFFADGGCSGGAYDVIICD 1317

Qy       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380

Db       1318 ECHSQDATTILGIGTVLDQAETAGARLVVLATAIPPGSVTTPHPNIEEVALPSEGEIPFY 1377

Qy       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440

Db       1378 GRAIPLVLIKGGRHLIFANQKKKAKETAKKMNKPGRKAVAYYRGLDVAVIPATGDVVVCS 1437

Qy       1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500

Db       1438 TDALMTGFTGDFDSVIDCNSAVTQTVDFSLDPTFTIETTTVPQDAVSRSQRRGRTGRGRH 1497

Qy       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560

Db       1498 GIYRYVSSGERPSGIFDSVVLCECYDAGCAWYDLTPAETTVRLRAYLNTPGLPVCQEHLE 1557

Qy       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620

Db       1558 FWEGVFTGLTNIDAHMLSQAKQGGENFPYLVAYQATVCVRAKAPPPSWDTMWKCMICLKP 1617

Qy       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680

Db       1618 TLTGPTPLLYRLGAVQNEITLTHPITKYIMACMSADLEVITSTWVLVGGVVAALAAYCLT 1677

Qy       1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740

Db       1678 VGSVAIVGRIILSGRPAITPDREVLYQQFDEMEECSASLPYVDEARAIAGQFKEKVLGLI 1737

Qy       1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800

Db       1738 GTAGQKAETLKPAATSMWSKAEQFWAKHMWNFVSGIQYLAGLSTLPGNPAVATLMSFTAA 1797

Qy       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860

Db       1798 VTSPLTTHQTLLFNILGGWVASQIAPPTAATAFVVSGMAGAAVGNIGLGRVLIDILAGYG 1857

Qy       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920

Db       1858 TGVAGALVAFKIMCGERPTAEELVNLLPSILCPGALVVGVICAAVLRRHIGPGEGAVQWM 1917

Qy       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980

Db       1918 NRLIAFASRGNHGSPTHYVPETDASAKVTQLLSSLTVTSLLKRLHTWIGEDYSTPCDGTW 1977

Qy       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040

Db       1978 LRAIWDWVCTALTDFKAWLQAKLLPQLPGVPFFSCQKGYKGVWRGDGVNSTKCPCGATIS 2037

Qy       2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100

Db       2038 GHVKNGTMRIVGPKLCSNTWQGTFPINATTTGPSVPAPAPNYKFALWRVGAADYAEVRRV 2097

Qy       2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160

Db       2098 GDYHYITGVTQDNLKCPCQVPSPEFFTELDGVRIHRFAPPCNPLLREEVTFSVGLHSYVV 2157

Qy       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220

Db       2158 GSQLPCEPEPDVTVLTSMLSDPAHITAETAKRRLNRGSPPSLANSSASQLSAPSLKATCT 2217

Qy       2221 THGKAYDVDMVDANLF----MGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYML 2276

Db       2218 IQGHHPDADLIKANLLWRQCMGGNITRVEAENKVEILDCFKPLKEEEDDREISVSADCFK 2277

Qy       2277 PKKRFPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRR-RRTVGL 2335

Db       2278 KGPAFPPALPVWARPGYDPPLLETWKRPDYDPPQVWGCPIPPAGPPPVPLPRRKRKPMEL 2337

Qy       2336 SEDSIGDALQQLAIKSFGQPPPS---GDSGLSTGAGAADSGSQTPPDELALSETGSISSM 2392

Db       2338 SDSTVSQVMADLADARFKVDTPSIEGQDSALGT-SSQHDSGPEEKRDD--NSDAASYSSM 2394
```

```
Qy       2393 PPLEGELGDPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGAL 2452
Db       2395 PPLEGEPGDPDL-------------------SSGSWSTVSGE-DNVVCCSMSYTWTGAL 2433

Qy       2453 ITPCSPEEEKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKD 2512
Db       2434 ITPCSAEEEKLPINPLSNTLLRHHNLVYSTSSRSAGLRQKKVTFDRLQVLDDHYREVVDE 2493

Qy       2513 IKLAASKVTARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDS 2572
Db       2494 MKRLASKVKARLLPLEEACGLTPPHSARSKYGYGAKEVRSLDKKALKHIEGVWQDLLDDS 2553

Qy       2573 ETPIPTTIMAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGAS 2632
Db       2554 DTPLPTTIMAKNEVFAVEPSKGGKKPARLIVYPDLGVRVCEKRALYDVAQKLPTALMGPS 2613

Qy       2633 YGFQYSPAQRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAH 2692
Db       2614 YGFQYSPAQRVDFLLKAWKSKKIPMAFSYDTRCFDSTITEHDIMTEESIYQSCDLQPEAR 2673

Qy       2693 TAIHSLTERLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIA 2752
Db       2674 VAIRSLTQRLYCGGPMYNSKGQQCGYRRCRASGVFTTSMGNTMTCYIKALASCRAAKLRD 2733

Qy       2753 PTMLVCGDDLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNV 2812
Db       2734 CTLLVCGDDLVAICESQGTHEDEASLRAFTEAMTRYSAPPGDPPVPAYDLELVTSCSSNV 2793

Qy       2813 SVALGPQGRRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFF 2872
Db       2794 SVARDASGNRIYYLTRDPQVPLAKAAWETAKHSPVNSWLGNIIMYAPTLWARIVLMTHFF 2853

Qy       2873 SILMAQDTLDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALR 2932
Db       2854 SVLQSQEQLEKTLAFEMYGSVYSVTPLDLPAIIQRLHGLSAFSLHSYSPSEINRVASCLR 2913

Qy       2933 KLGAPPLRAWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSW 2992
Db       2914 KLGVPPLRAWRHRARAVRAKLIAQGGRAAICGIYLFNWAVKTKRKLTPLADADRLDLSSW 2973

Qy       2993 FTVGAGGGDIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
Db       2974 FTVGAGGGDIYHSMSRARPRNLLLCLLLLSVGVGIFLLPAR 3014


RESULT 12
JQ0881
genome polyprotein - hepatitis C virus (strain J6) (fragments)
N;Contains: NS5 protein
C;Species: hepatitis C virus
C;Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 09-Jul-2004
C;Accession: JQ0881
R;Okamoto, H.
submitted to JIPID, January 1991
A;Description: The 5'-terminal and 3'-terminal sequences of the genomic RNA of hepatitis
A;Reference number: JQ0879
A;Accession: JQ0881
A;Molecule type: genomic RNA
A;Residues: 1-874 <OKA>
A;Cross-references: UNIPROT:Q7LZY5
A;Experimental source: strain J6
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: polyprotein
F;510-874/Product: NS5 protein (fragment) #status predicted <NS5>

  Query Match            21.7%;  Score 3490.5;  DB 2;  Length 874;
  Best Local Similarity  28.1%;  Pred. No. 1.7e-191;
  Matches  853;  Conservative    8;  Mismatches   13;  Indels 2159;  Gaps    1;

Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
Db         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
Db        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
Db        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNTSRCWI 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
Db        241 PVSPNVAVQQPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
Db        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIIGGAHWGVMFGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
Db        361 YFSMQGAWAKVVVILLLAAGVDAQTHTVGGSTAHNARTLTGMFSLGARQKIQLINTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
Db        421 HINRTALNCNDSLHTGFLASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
Db        481 PEDMRPYCWHYPPRQCGVVSASSVCGPVY---------------------------- 509

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
Db        510 ------------------------------------------------------------ 509

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
Db        510 ------------------------------------------------------------ 509

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
Db        510 ------------------------------------------------------------ 509

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
Db        510 ------------------------------------------------------------ 509

Qy        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
Db        510 ------------------------------------------------------------ 509

Qy        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
Db        510 ------------------------------------------------------------ 509

Qy        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
Db        510 ------------------------------------------------------------ 509

Qy        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
Db        510 ------------------------------------------------------------ 509

Qy       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080
Db        510 ------------------------------------------------------------ 509

Qy       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
Db        510 ------------------------------------------------------------ 509

Qy       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200
```

```
Db        510 ------------------------------------------------------------ 509
Qy       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260
Db        510 ------------------------------------------------------------ 509
Qy       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320
Db        510 ------------------------------------------------------------ 509
Qy       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380
Db        510 ------------------------------------------------------------ 509
Qy       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440
Db        510 ------------------------------------------------------------ 509
Qy       1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500
Db        510 ------------------------------------------------------------ 509
Qy       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560
Db        510 ------------------------------------------------------------ 509
Qy       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
Db        510 ------------------------------------------------------------ 509
Qy       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680
Db        510 ------------------------------------------------------------ 509
Qy       1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740
Db        510 ------------------------------------------------------------ 509
Qy       1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
Db        510 ------------------------------------------------------------ 509
Qy       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860
Db        510 ------------------------------------------------------------ 509
Qy       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920
Db        510 ------------------------------------------------------------ 509
Qy       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980
Db        510 ------------------------------------------------------------ 509
Qy       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040
Db        510 ------------------------------------------------------------ 509
Qy       2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100
Db        510 ------------------------------------------------------------ 509
Qy       2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160
Db        510 ------------------------------------------------------------ 509
Qy       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220
Db        510 ------------------------------------------------------------ 509
Qy       2221 THGKAYDVDMVDANLFMGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYMLPKKR 2280
```

```
Db        510 ------------------------------------------------------------ 509
Qy       2281 FPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLSEDSI 2340
Db        510 ------------------------------------------------------------ 509
Qy       2341 GDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTPPDELALSETGSISSMPPLEGELG 2400
Db        510 ------------------------------------------------------------ 509
Qy       2401 DPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460
Db        510 ------------------------------------------------------------ 509
Qy       2461 EKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIKLAASKV 2520
Db        510 ------------------------------------------------------------ 509
Qy       2521 TARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSETPIPTTI 2580
Db        510 ------------------------------------------------------------ 509
Qy       2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPA 2640
Db        510 ------------------------------------------------------------ 509
Qy       2641 QRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTE 2700
Db        510 ----------------------------TVTERDIRTEESIYRACSLPEEAHTAIHSLTE 541
Qy       2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPTMLVCGD 2760
Db        542 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPTMLVCGD 601
Qy       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820
Db        602 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 661
Qy       2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDT 2880
Db        662 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDT 721
Qy       2881 LDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2940
Db        722 LDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 781
Qy       2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000
Db        782 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 841
Qy       3001 DIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
Db        842 DIYHSVSRARPRLLLLGLLLLFVGVGLFLLPAR 874
```

RESULT 13
PC2219
polypeptide - hepatitis C virus (type 5a) (fragments)
N;Contains: core protein; E1 (carboxyl end); E2/NS1 (amino end); NS3 protein; NS4A prote
C;Species: hepatitis C virus
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Jul-2004
C;Accession: PC2219
R;Stuyver, L.; Arnhem, W.V.; Wyseur, A.; Maertens, G.
Biochem. Biophys. Res. Commun. 202, 1308-1314, 1994
A;Title: Cloning and phylogenetic analysis of the core, E2, and NS3/NS4 regions of the h
A;Reference number: PC2219; MUID:94338342; PMID:7520237
A;Accession: PC2219
A;Molecule type: mRNA
A;Residues: 1-876 <STU>
A;Cross-references: UNIPROT:Q81242; GB:L29577; GB:L29578; GB:L29579
A;Experimental source: serum
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: glycoprotein

```
F;1-191/Product: core #status predicted <COE>
F;68-78/Region: variable
F;192-247/Product: E1 (carboxyl end) #status predicted <ERE>
F;248-411/Product: E2/NS1 (amino end) #status predicted <ENR>
F;248-338/Region: E2
F;339-411/Region: NS1 (amino end)
F;412-783/Product: NS3 #status predicted <NSR>
F;784-837/Product: NS4A #status predicted <NSA>
F;838-876/Product: NS4B #status predicted <NSB>
F;281,287,294,312,340/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match            20.2%;  Score 3263;  DB 2;  Length 876;
  Best Local Similarity  38.1%;  Pred. No. 1.8e-178;
  Matches  667;  Conservative   86;  Mismatches  121;  Indels  878;  Gaps    3;

Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              ||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||
Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRMGVRATRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              ||||||| |: ||:|||:|||||||| |||||||||||||||||:||||||| :|||:|
Db         61 RRQPIPKARQPTGRSWGQPGYPWPLYANEGLGWAGWLLSPRGSRPNWGPNDPRRKSRNLG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
              ||||||||||||||||:|| |:|||||||||||||||:|||||||||||||:||
Db        121 KVIDTLTCGFADLMGYIPLVGGPIGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFILA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
              ||||:| | ||
Db        181 LLSCLTVPASA---------------------------------------------------- 191

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300

Db        192 ------------------------------------------------------------ 191

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
                                          ||  :::|  :|:|:|:||||:|:||||:|  |
Db        192 --------------------------PTTALLVAQLLRIPQVVIDIIAGSHWGVLFAAA 224

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
              |::    | |||::| | |||||  |   |||:|||| : |    | :||:||:||||||
Db        225 YYASVANWTKVVLVLFLFAGVDATTQISGGSSAQTTYGIASFITRGAQQKLQLINTNGSW 284

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
              |||||||||||| ||||| ||| | |||||||:||::||::  |  ||| : |  |::
Db        285 HINRTALNCNDSLQTGFIAGLFYYHKFNSSGCPDRMASCRALAFFDQGWGTISYA-NISG 343

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
              | | :|||||||| |||| |: |||||||||||||||||  | ||| || | || ||
Db        344 PSDDKPYCWHYPPRPCGVVPAQEVCGPVYCFTPSPVVVGTTDSKGHPTYNWGSNVTDFFL 403

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
              :|:||||
Db        404 MNNTRPP---------------------------------------------------- 410

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660

Db        411 ------------------------------------------------------------ 410

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720

Db        411 ------------------------------------------------------------ 410

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780

Db        411 ------------------------------------------------------------ 410

Qy        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840

Db        411 ------------------------------------------------------------ 410

Qy        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900

Db        411 ------------------------------------------------------------ 410

Qy        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960

Db        411 ------------------------------------------------------------ 410

Qy        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020

Db        411 ------------------------------------------------------------ 410

Qy       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080

Db        411 ------------------------------------------------------------ 410

Qy       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140

Db        411 ------------------------------------------------------------ 410

Qy       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200

Db        411 ------------------------------------------------------------ 410

Qy       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260

Db        411 ------------------------------------------------------------ 410

Qy       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320
                                          :|||| ||||||||||||||:|||||:||||
Db        411 ----------------------------ITTGASITYSTYGKFLADGGCSGGAYDVIICD 442

Qy       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380
              |||: |:|||||||||||||||| || |||||||||||||||||||||| ||||:|||
Db        443 ECHSQDATTILGIGTVLDQAETAGARLVVLATATPPGSVTTPHPNIEEVALPQEGEVPFY 502

Qy       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440
              ||||||:|||||||||||||||||||  |  :|:|:|||||||||:|||| ||||| :
Db        503 GRAIPLAFIKGGRHLIFCHSKKKCDELAKQLTSLGVNAVAYYRGLDVAVIPTAGDVVVCS 562

Qy       1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500
              |||||||:|||||||||| |||| |||||||||| |  ||||||||||||||||||
Db        563 TDALMTGFTGDFDSVIDCNSAVTQTVDFSLDPTFTIETTTVPQDAVSRSQRRGRTGRGRH 622

Qy       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560
              ||||||| ||| | |||||||||||||| |||:|||:||||||||| ||||||||||||
Db        623 GIYRYVSAGERPSDMFDSVVLCECYDAGCAWYDLTPAETTVRLRAYINTPGLPVCQDHLE 682

Qy       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
              ||| ||||||:|||| |||||| |||| || ||||||| |||||||||| ||||: ||||
Db        683 FWEGVFTGLTNIDAHMLSQTKQGGENFPYLVAYQATVCVRAKAPPPSWDTMWKCMLRLKP 742

Qy       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680
              || |||||||||| | ||:|||||:|||| || |||||:|||||| |||:||:|||||
Db        743 TLTGPTPLLYRLGPVQNEITLTHPITKYIMACMSADLEVITSTWVLVGGVVAALAAYCLT 802

Qy       1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740
              | | |:||: :: :  : ||:| ||: |||||||::    ::| : ||   | |: | :
Db        803 VGSVAIVGRIILSGKPAIIPDREALYQQFDEMEECSASLPYMDETRAIAGQFKEKVLGFI 862

Qy       1741 QQASKQAQDIQP 1752
                  ::|: ::|
Db        863 STTGQKAETLKP 874


RESULT 14
S19875
genome polyprotein - hepatitis C virus (isolate JK3) (fragment)
N;Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein
C;Species: hepatitis C virus
A;Variety: isolate JK3
C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 09-Jul-2004
C;Accession: S19875
R;Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
submitted to the EMBL Data Library, September 1991
```

A;Description: Sequence analysis of putative structural regions of Hepatitis C Virus isc
A;Reference number: S18029
A;Accession: S19875
A;Molecule type: genomic RNA
A;Residues: 1-782 <HON>
A;Cross-references: UNIPROT:Q68951; EMBL:X61592; NID:g59482; PIDN:CAA43789.1; PID:g59483
A;Experimental source: isolate JK3
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural
F;1-191/Product: core protein #status predicted <MAT1>
F;192-383/Product: envelope protein 1 #status predicted <MAT2>
F;384-733/Product: NS1/E2 protein #status predicted <MAT3>
F;734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 20.0%; Score 3231; DB 2; Length 782;
Best Local Similarity 70.7%; Pred. No. 1e-176;
Matches 556; Conservative 104; Mismatches 122; Indels 4; Gaps 2;

```
Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              |||||||||:||||| ||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSTNPKPQRQTKRNTIRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              ||||||| |:  |::| :|||||||||||||:|| |||||||||||||| ||| ||||:|
Db         61 RRQPIPKARQPEGRAWAQPGYPWPLYGNEGIGWTGWLLSPRGSRPSWGPTDPRRRSRNLG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
              ||||||||||||||||||:||||||||||||||||||||||:||||||||||||||||||
Db        121 KVIDTLTCGFADLMGYIPLVGAPLGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
              ||||:| | || ||:|:|  | |||||:| || ::    ::| |||||| : || |:||:
Db        181 LLSCLTVPASAYEVRNVSGAYHVTNDCSNTSIVYEAADMIMHTPGCVPCVRDGNISRCWV 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
              ::  :| :     |  :| |:|::| :| ||||:|||||| | | :|:|  ||: :
Db        241 ALTLTLAARNSSIPTTTIRRHVDLLVGAAALCSAMYVGDLCGSVFLVSQLFTFSPRRYET 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
              |||||||||| :|||||||||||||||  ::::  :|:|: ::|:::|||||: |||
Db        301 VQDCNCSIYPGKLTGHRMAWDMMMNWSPTTALVVSQLLRIPQAVVDMVAGAHWGVLAGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
              |:|| | ||||::::|| |||| :|    |    :||  |||||  || | |||||:||||
Db        361 YYSMVGNWAKVLIVMLLFAGVDGQTRVTGAQVGRTTSSLTSLFTPGPSQNIQLVNSNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
              |||||||:|||||  |||:|:|||||  ||:|||||||::||||: |  |||  : :   | |
Db        421 HINRTALSCNDSLKTGFLAALFYTHKFNASGCPERMASCRSIDTFDQGWGPITHV--VPN 478

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
              | :|||||| || ||:|  |  ||||||||||||||||||| |||||||||||||| |
Db        479 TTDQKPYCWHYAPRPCGIVPASQVCGPVYCFTPSPVVVGTTDRSGAPTYTWGENETDVLL 538

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
              ||:|||| |:|||||||||:|:|||| |||       |    |:|||||||||: || |
Db        539 LNNTRPPHGNWFGCTWMNSTGFTKTCGGPPCNIGGAGN--NTLICPTDCFRKHPEATYTK 596

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
              ||||||||||::|||||||||||| |:::|||||||||||||  ||||:|||:||:| |
Db        597 CGSGPWLTPRCMVDYPYRLWHYPCTANFSVFKIRMYVGGVEHRFNAACNWTRGERCDLSD 656

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
              |||::||||| ||||| :||||:: |||||||:||||||||::||:  |:   :|||
Db        657 RDRAELSPLLLSTTEWQVLPCSFTTLPALSTGLIHLHQNIVDVQYLYGIGSAVVSVVVRW 716

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
              |:::|||||||||||||::|: ||||||| ||:|:||| |    |  :|::|| |||
Db        717 EYILLLFLLLADARVCACLWMMLLIAQAEAALENLVVLNAASVAGTRGIPFFLVFFCAAW 776

Qy        781 YIKGRV 786
              |||||:
Db        777 YIKGRL 782
```

RESULT 15
S19876
genome polyprotein - hepatitis C virus (isolate JK5) (fragment)
N;Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein
C;Species: hepatitis C virus
A;Variety: isolate JK5
C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 09-Jul-2004
C;Accession: S19876
R;Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
submitted to the EMBL Data Library, September 1991
A;Description: Sequence analysis of putative structural regions of Hepatitis C Virus isc
A;Reference number: S18029
A;Accession: S19876
A;Molecule type: genomic RNA
A;Residues: 1-782 <HON>
A;Cross-references: UNIPROT:Q68953; EMBL:X61595; NID:g59486; PIDN:CAA43792.1; PID:g59487
A;Experimental source: isolate JK5
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural
F;1-191/Product: core protein #status predicted <MAT1>
F;192-383/Product: envelope protein 1 #status predicted <MAT2>
F;384-733/Product: NS1/E2 protein #status predicted <MAT3>
F;734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 19.9%; Score 3214; DB 2; Length 782;
Best Local Similarity 70.6%; Pred. No. 9.5e-176;
Matches 555; Conservative 102; Mismatches 125; Indels 4; Gaps 2;

```
Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLFPRRGPRLGVRATRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              ||||||| |:  |::| :|||||||||||||||||||||||||||||| ||| ||||:|
Db         61 RRQPIPKARQPEGRAWAQPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPTDPRRRSRNLG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
              ||||||||||| |||||:|||||| |||||||||||||||:||||||||||||||||||
Db        121 KVIDTLTCGFADPMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
              ||||:| | || ||:|:|  | |||||:| || ::    ::| ||||||   ||:|:||:
Db        181 LLSCLTIPASAYEVRNVSGLYHVTNDCSNSSIVYEAADLIMHTPGCVPCVLEGNSSRCWV 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
              ::| :| :     |  :| |:|::| :|  |||:|||||| | | :|:|  ||: :
Db        241 ALTPTLAARNNSIPTTTIRRHVDLLVGAAAFCSAMYVGDLCGSVFLVSQLFTFSPRRYET 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
              |||||||:||| ::|||||||||||||  ::::  :|:|: ::|:::|||||: |||
Db        301 VQDCNCSLYPGHVSGHRMAWDMMMNWSPTIALVVSQLLRIPQAVVDMVAGAHWGVLAGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
              |:|| | ||||:::|  |||| :|    |    :||  |||||  || | |||||:||||
Db        361 YYSMVGNWAKVLIVMLRFAGVDGQTRVTGAQVGRTTSSLTSLFTPGPSQNIQLVNSNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
              |||||||:|||| |||:|:|||| ||:|||||||::||||: |  |||  : :   | |
Db        421 HINRTALSCNDSLKTGFLAALFYTHKFNASGCPERMASCRSIDTFDQGWGPITHV--VPN 478

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
              | :|||||| || ||:| |  ||||||||||||||||||| |||||| || |||||  :
Db        479 TTDQKPYCWHYAPRPCGIVPASQVCGPVYCFTPSPVVVGTTDRFGAPTYNWGANETDVLV 538

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
              ||:|||| |:|||||||||:|:|||| |||       |    | ||||||||: || |
Db        539 LNNTRPPQGNWFGCTWMNSTGFTKTCGGPPCNIGGAGN--NTLTCPTDCFRKHPEATYTK 596

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
              |||||||||||:: |||||||||||||:|||||||||||||||| || |:||||||:|||
Db        597 CGSGPWLTPRCMVHYPYRLWHYPCTVNPTIFKIRMYVGGVEHRLKAASNWTRGDRCDLED 656
```

```
Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
              ||||:||||: :|||| |||||:: |||||||:|||||||||::||:  |:   :::|
Db        657 RDRSELSPLIRATTEWQILPCSFTTLPALSTGLIHLHQNIVDVQYLYGVGSAVVSIVIKW 716

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
              |:|:|||||||||||||||::|: ||||||| ||:|:||| |  :| | |::|| |||
Db        717 EYVLLLFLLLADARVCACLWMMLLIAQAEAALENLVVLNAASVAGAHGILSFLVFFCAAW 776

Qy        781 YIKGRV 786
              |||||:
Db        777 YIKGRL 782
```

Search completed: October 28, 2005, 16:06:03
Job time : 97 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2005, 15:36:16 ; Search time 328 Seconds
(without alignments)
4735.171 Million cell updates/sec

Title: US-09-980-559-2
Perfect score: 16120
Sequence: 1 MSTNPKPQRKTKRNTNRRPQ.....LLFGLLLLFVGVGLFLLPAR 3033

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	16120	100.0	3033	2	Q9QF35	Q9qf35 hepatitis c
2	15830	98.2	3033	1	POLG_HCVJ6	P26660 h genome po
3	15426	95.7	3033	2	Q9Q9A7	Q9q9a7 hepatitis c
4	15394	95.5	3033	2	Q99IB6	Q99ib6 hepatitis c
5	15356	95.3	3033	2	Q9IZA4	Q9iza4 hepatitis c
6	15355	95.3	3033	2	Q9IZA5	Q9iza5 hepatitis c
7	15354	95.2	3033	2	Q99IB7	Q99ib7 hepatitis c
8	15350	95.2	3033	2	Q99IB3	Q99ib3 hepatitis c
9	15347	95.2	3033	2	Q9IZA6	Q9iza6 hepatitis c
10	15324.5	95.1	3032	2	Q99IB2	Q99ib2 hepatitis c
11	15322	95.0	3033	2	Q9IZA3	Q9iza3 hepatitis c
12	15321	95.0	3033	2	Q9IZA2	Q9iza2 hepatitis c
13	15311	95.0	3033	2	Q9Q9B0	Q9q9b0 hepatitis c
14	15266	94.7	3033	2	Q9Q9A8	Q9q9a8 hepatitis c
15	15249.5	94.6	3032	2	Q9Q9A9	Q9q9a9 hepatitis c
16	15207	94.3	3033	2	Q99IB5	Q99ib5 hepatitis c
17	15115.5	93.8	3032	2	Q99IB4	Q99ib4 hepatitis c
18	14943	92.7	3033	2	Q99IB8	Q99ib8 hepatitis c
19	14356	89.1	3037	2	Q68749	Q68749 hepatitis c
20	14185	88.0	3033	2	Q9QAX1	Q9qax1 hepatitis c
21	14014	86.9	3033	2	Q7T7J1	Q7t7j1 hepatitis c
22	14013	86.9	3033	2	Q7T7H3	Q7t7h3 hepatitis c
23	14007	86.9	3033	2	Q7T7I1	Q7t7i1 hepatitis c
24	14005	86.9	3033	2	Q7T7H6	Q7t7h6 hepatitis c
25	14003	86.9	3033	2	Q7T7H5	Q7t7h5 hepatitis c
26	13998	86.8	3033	2	Q7T7I2	Q7t7i2 hepatitis c
27	13990	86.8	3033	2	Q7T7H4	Q7t7h4 hepatitis c
28	13987	86.8	3033	2	Q7T7H7	Q7t7h7 hepatitis c
29	13987	86.8	3033	2	Q7T7J2	Q7t7j2 hepatitis c
30	13976	86.7	3033	2	Q7T7H8	Q7t7h8 hepatitis c
31	13971	86.7	3033	1	POLG_HCVJ8	P26661 h genome po
32	13965	86.6	3033	2	Q7T7H9	Q7t7h9 hepatitis c
33	13961	86.6	3033	2	Q7T7I9	Q7t7i9 hepatitis c
34	13960	86.6	3033	2	Q7T7I0	Q7t7i0 hepatitis c
35	13939	86.5	3033	2	Q7T7I6	Q7t7i6 hepatitis c
36	13937	86.5	3033	2	Q7T7I3	Q7t7i3 hepatitis c
37	13919	86.3	3033	2	Q7T7J0	Q7t7j0 hepatitis c
38	13917	86.3	3033	2	Q7T7I8	Q7t7i8 hepatitis c
39	13914	86.3	3033	2	Q7T7I7	Q7t7i7 hepatitis c
40	13912	86.3	3033	2	Q7T7I4	Q7t7i4 hepatitis c
41	13885	86.1	3033	2	Q9DHD6	Q9dhd6 hepatitis c
42	13884	86.1	3033	2	Q7T7I5	Q7t7i5 hepatitis c
43	13884	86.1	3033	2	Q9IZA1	Q9iza1 hepatitis c
44	13049	80.9	3015	2	Q9PWU9	Q9pwu9 hepatitis c
45	12925	80.2	3015	2	Q9PWX5	Q9pwx5 hepatitis c

ALIGNMENTS

```
RESULT 1
Q9QF35
ID   Q9QF35      PRELIMINARY;      PRT;  3033 AA.
AC   Q9QF35;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HC-J6CH;
RX   MEDLINE=99420396; PubMed=10489358; DOI=10.1006/viro.1999.9889;
RA   Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;
RT   "Hepatitis C virus: an infectious molecular clone of a second major
RT   genotype (2a) and lack of viability of intertypic 1a and 2a
RT   chimeras.";
RL   Virology 262:250-263(1999).
DR   EMBL; AF177036; AAF01178.1; -.
DR   HSSP; Q8JYS1; 1CWX.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR011545; DEAD/DEAH_N.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR004109; Peptidase_S29.
DR   InterPro; IPR009003; Pept_Ser_Cys.
DR   InterPro; IPR002518; Pept_U39_HCV_NS2.
DR   InterPro; IPR002129; Pyridoxal_deC.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
```

```
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
DR   PROSITE; PS00392; DDC_GAD_HDC_YDC; UNKNOWN_1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; Transmembrane.
SQ   SEQUENCE   3033 AA;  329264 MW;  6D894082FFF885C7 CRC64;

  Query Match           100.0%;  Score 16120;  DB 2;  Length 3033;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 3033;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780

Qy        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840

Qy        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900

Qy        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960

Qy        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020

Qy       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080

Qy       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140

Qy       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200

Qy       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260

Qy       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320

Qy       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380

Qy       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440

Qy       1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500

Qy       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560

Qy       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620

Qy       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680

Qy       1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740

Qy       1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800

Qy       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860

Qy       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920

Qy       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980

Qy       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040

Qy       2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100

Qy       2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160

Qy       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220

Qy       2221 THGKAYDVDMVDANLFMGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYMLPKKR 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 THGKAYDVDMVDANLFMGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYMLPKKR 2280

Qy       2281 FPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLSEDSI 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 FPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLSEDSI 2340

Qy       2341 GDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTPPDELALSETGSISSMPPLEGELG 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTPPDELALSETGSISSMPPLEGELG 2400

Qy       2401 DPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 DPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460

Qy       2461 EKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIKLAASKV 2520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 EKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIKLAASKV 2520

Qy       2521 TARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSETPIPTTI 2580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2521 TARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSETPIPTTI 2580

Qy       2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPA 2640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPA 2640

Qy       2641 QRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTE 2700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2641 QRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTE 2700

Qy       2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPTMLVCGD 2760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPTMLVCGD 2760

Qy       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820

Qy       2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDT 2880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDT 2880

Qy       2881 LDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2881 LDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2940

Qy       2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000

Qy       3001 DIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
              ||||||||||||||||||||||||||||||||
Db       3001 DIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033


RESULT 2
POLG_HCVJ6
ID   POLG_HCVJ6     STANDARD;      PRT;  3033 AA.
AC   P26660;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   25-OCT-2004 (Rel. 45, Last annotation update)
DE   Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE   Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE   (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE   (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE   (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE   NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE   NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS   Hepatitis C virus (isolate HC-J6) (HCV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11113;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92044440; PubMed=1658196;
RA   Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Lizuka H., Machida A.,
RA   Miyakawa Y., Mayumi M.;
RT   "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated
RT   from a human carrier: comparison with reported isolates for conserved
RT   and divergent regions.";
RL   J. Gen. Virol. 72:2697-2704(1991).
CC   -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC       hydrophobic, suggesting a possible membrane-related function. NS3
CC       and NS5 may play a role in the viral RNA replication.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC       precursor polyprotein, commonly with Asp or Glu in the P6
CC       position, Cys or Thr in P1 and Ser or Ala in P1'.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC       lipoprotein envelope. The envelope consists of two proteins:
CC       protein M and glycoprotein E. The nucleocapsid is a complex of
CC       protein C and mRNA.
CC   -!- SIMILARITY: Contains 1 peptidase S29 domain.
CC   -!- SIMILARITY: Contains 1 peptidase U39 domain.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; D00944; BAA00792.1; -.
DR   PIR; JQ1303; JQ1303.
DR   HSSP; P27958; 1HEI.
DR   MEROPS; S29.001; -.
DR   MEROPS; U39.001; -.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR009003; Pept_Ser_Cys.
DR   InterPro; IPR002518; Pept_U39_HCV_NS2.
```

```
DR   InterPro; IPR004109; Peptidase_S29.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; Helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   ATP-binding; Coat protein; Core protein; Envelope protein;
KW   Glycoprotein; Helicase; Hydrolase; Nonstructural protein; Polyprotein;
KW   RNA-directed RNA polymerase; Serine protease; Transferase;
KW   Transmembrane.
FT   INIT_MET      1      1       Removed from capsid protein C by the
FT                                cellular aminopeptidase.
FT   CHAIN         1    115       Capsid protein C (Potential).
FT   CHAIN       116    191       Matrix protein (Potential).
FT   CHAIN       192    383       Major envelope protein E (Potential).
FT   CHAIN       384    733       Nonstructural protein NS1 (Potential).
FT   CHAIN       734   1010       Nonstructural protein NS2 (Potential).
FT   CHAIN      1011   1619       Protease/helicase NS3 (Potential).
FT   CHAIN      1620   1866       Nonstructural protein NS4A (Potential).
FT   CHAIN      1867   2017       Nonstructural protein NS4B (Potential).
FT   CHAIN      2018   3033       RNA-directed RNA polymerase (Potential).
FT   TRANSMEM    347    369       Potential.
FT   ACT_SITE   1087   1087       Charge relay system (By similarity).
FT   ACT_SITE   1111   1111       Charge relay system (By similarity).
FT   ACT_SITE   1169   1169       Charge relay system (By similarity).
FT   NP_BIND    1234   1241       ATP (Potential).
FT   SITE       1320   1323       DECH box.
FT   CARBOHYD    196    196       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    209    209       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    234    234       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    305    305       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    417    417       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    423    423       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    430    430       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    448    448       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    477    477       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    534    534       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    542    542       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    558    558       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    578    578       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    627    627       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    649    649       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD   1091   1091       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD   2038   2038       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD   2811   2811       N-linked (GlcNAc. . .) (Potential).
SQ   SEQUENCE   3033 AA;  329165 MW;  F957F5C1A273BE9E CRC64;

  Query Match             98.2%;  Score 15830;  DB 1;  Length 3033;
  Best Local Similarity   97.8%;  Pred. No. 0;
  Matches 2967;  Conservative   30;  Mismatches   36;  Indels    0;  Gaps    0;

Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |:|||
Db        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNTSRCWI 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
              |||||||||:||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PVSPNVAVQQPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIIGGAHWGVMFGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
              |||||||||||||||||||||:||||||| |     || :| :| ||||||:||||||
Db        361 YFSMQGAWAKVVVILLLAAGVDAQTHTVGGSTAHNARTLTGMFSLGARQKIQLINTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
              |||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db        421 HINRTALNCNDSLHTGFLASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
              |||||||||||||||||| :|||||||||||||||||||||||||||||||||||||||
Db        481 PEDMRPYCWHYPPRQCGVVSASSVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
              ||||||| |||||||||:|||||||| |||||| |||||||||||||||:|
Db        541 LNSTRPPQGSWFGCTWMNSTGYTKTCGAPPCRIRADFNASMDLLCPTDCFRKHPDTTYIK 600

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
              ||||||||||||||||||:||||||||||||||||||||||||||||||||||||||||
Db        661 RDRSQLSPLLHSTTEWAILPCTYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
              |||:|||||||||||||||||||||||||||||:|||||||||||||||||||||||||
Db        721 EWVVLLFLLLADARVCACLWMLILLGQAEAALEKLVVLHAASAASCNGFLYFVIFFVAAW 780

Qy        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
              |||||||||||||||||| ||||||||||||||||||||||:||||||||||
Db        781 YIKGRVVPLATYSLTGLWSFGLLLLALPQQAYAYDASVHGQIGAALLVLITLFTLTPGYK 840

Qy        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
              |||||||||||||| |||||||||||||||||||||||||| |||||||||||||||
Db        841 TLLSRFLWWLCYLLTLAEAMVQEWAPPMQVRGGRDGIIWAVAIFCPGVVFDITKWLLAVL 900

Qy        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
              |||||||||||||||||||||||||||||| |||||||||| |||||||||||||||||
Db        901 GPAYLLKGALTRVPYFVRAHALLRMCTMVRHLAGGRYVQMVLLALGRWTGTYIYDHLTPM 960

Qy        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
              |||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SDWAANGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020

Qy       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Db       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTTISG 1080

Qy       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140

Qy       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200

Qy       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260

Qy       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320

Qy       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380

Qy       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440
              ||||||||||||||||||||||||||||||||||||||:|||||||||||||||||||||
Db       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNAVAYYRGLDVSVIPTQGDVVVVA 1440

Qy       1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500
              |||||||:|||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       1441 TDALMTGFTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQTVPQDAVSRSQRRGRTGRGRL 1500

Qy       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560
              ||||||||||||||||||||||||||||||||||:|||||||||||||||||||||||||
Db       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPAETTVRLRAYFNTPGLPVCQDHLE 1560

Qy       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620

Qy       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680

Qy       1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740
              ||||||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TGCVCIIGRLHVNQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740

Qy       1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 QQASKQAQDIQPAVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800

Qy       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860

Qy       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920

Qy       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCSGSW 1980

Qy       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040
              ||||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFISCQKGYKGVWAGTGIMTTRCPCGANIS 2040

Qy       2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100
              ||||||||||||||||||||||||||||||||||||||||||||:||||||||||||||||
Db       2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKIAIWRVAASEYAEVTQH 2100

Qy       2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db       2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPIPKPFFRDEVSFCVGLNSFVV 2160

Qy       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 GSQLPCDPEPDTDVLTSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220

Qy       2221 THGKAYDVDMVDANLFMGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYMLPKKR 2280
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       2221 THGKAYDVDMVDANLFMGGDVTRIESESKVVVLDSLDPMVEERSDLEPSIPSEYMLPKKR 2280

Qy       2281 FPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLSEDSI 2340
              ||||||||||||||||||||||||||||||||||||||:|||||||||||||||||| ||
Db       2281 FPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPKKTPTPPPRRRRTVGLSESSI 2340

Qy       2341 GDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTPPDELALSETGSISSMPPLEGELG 2400
              |||||||||||||||||||||||| ||||||:||||||||||||||||||||||||| |
Db       2341 ADALQQLAIKSFGQPPPSGDSGLSTGADAADSGSRTPPDELALSETGSISSMPPLEGEPG 2400

Qy       2401 DPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460
              |||||||||| |||||||  ||| ||||||||||||||||||||||||||||||||||||
Db       2401 DPDLEPEQVELQPPPQGGVVTPGSGSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460

Qy       2461 EKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIKLAASKV 2520
              ||||||||||||||||||||||:|||||||||||||| ||::|||||||||||||||||||
Db       2461 EKLPINPLSNSLLRYHNKVYCTTSKSASLRAKKVTFDRMQALDAHYDSVLKDIKLAASKV 2520

Qy       2521 TARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSETPIPTTI 2580
              ||||||:|||||||||||||||||||||||||||||||||||||||||||||||::|||||||
Db       2521 TARLLTLEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDTQTPIPTTI 2580

Qy       2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPA 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPA 2640

Qy       2641 QRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTE 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2641 QRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTE 2700

Qy       2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPTMLVCGD 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPTMLVCGD 2760

Qy       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820

Qy       2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDT 2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDT 2880

Qy       2881 LDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2881 LDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2940

Qy       2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000

Qy       3001 DIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
              ||||||||||||||| ||||||||||||||||
Db       3001 DIYHSVSRARPRLLLLGLLLLFVGVGLFLLPAR 3033
```

RESULT 3
Q9Q9A7
ID Q9Q9A7 PRELIMINARY; PRT; 3033 AA.
AC Q9Q9A7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21361470; PubMed=11468731; DOI=10.1002/jmv.1073;
RA Kurihara C., Ishiyama N., Nishiyama Y., Fukushi S., Kageyama T.,
RA Katayama K., Miura S.;
RT "Molecular characterization of hepatitis C virus genotype 2a from the
RT entire sequences of four isolates.";
RL J. Med. Virol. 64:466-475(2001).
DR EMBL; AF169005; AAF25613.1; -.
DR HSSP; Q8JYS1; 1CWX.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.

```
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR011545; DEAD/DEAH_N.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR004109; Peptidase_S29.
DR   InterPro; IPR009003; Pept_Ser_Cys.
DR   InterPro; IPR002518; Pept_U39_HCV_NS2.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; Transmembrane.
SQ   SEQUENCE   3033 AA;  329227 MW;  21492388CA0D5D8C CRC64;

  Query Match             95.7%;  Score 15426;  DB 2;  Length 3033;
  Best Local Similarity   94.8%;  Pred. No. 0;
  Matches 2874;  Conservative   72;  Mismatches   87;  Indels    0;  Gaps    0;

Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              ||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSTNPKPQRKTKRNTSRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPTDPRHRSRNVG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
              ||||||||||||||||||||||||||||||||||||||||||:||||||||||||||||
Db        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
              |||||| |||| :|||||  |||||||:|||||||||||||||||||||:|| |:|||
Db        181 LLSCITVPVSAVQVKNISDSYMVTNDCSNDSITWQLQAAVLHVPGCVPCEKMGNISRCWI 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
              |||||||||||||||| |||||||||||||||||||||||||||||||||||||||| |
Db        241 PVSPNVAVQRPGALTQGLRAHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHHF 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
              ||:|||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 VQECNCSIYPGAITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
              |||||||||||||||| ||||| | :: || |  |  |  ||  | :| |||:||||||
Db        361 YFSMQGAWAKVVVILLLTAGVDAHTRSIAGSVAHATSGLAGLFTSGAKQNIQLINTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
              ||||||||||||:||||||||| |||||||:|||| |:|||:||| |:|||||||||||
Db        421 HINRTALNCNDSLNTGFIASLFYTYRFNSSGCPERLSACRGIQAFRIGWGTLRYEDNVTN 480

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
              |||||||||||||:|||:|||::|||||||||||||||||||||| ||||||||||||:
Db        481 PEDMRPYCWHYPPKQCGIVSARSVCGPVYCFTPSPVVVGTTDRLGVPTYTWGENETDVFI 540

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
              ||||||| ||||||||||:|:|||||||||| ||||||| |||||||||||||| ||:|
Db        541 LNSTRPPGGSWFGCTWMNSTGFTKTCGAPPCRIRADFNASMDLLCPTDCFRKHPDATYIK 600

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
              ||||||||||||:|||||||||:|||||||||||||||||||||||||||||| |||||
Db        601 CGSGPWLTPRCLVDYPYRLWHYPCTINYTIFKIRMYVGGVEHRLTAACNFTRGDPCNLED 660

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
              ||||||||||||||||||||||||||||||||||||||||||||||:||||||||||||||
Db        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQYMYGLSPALTKYIVRW 720

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
              |||:|||||||||||||||||||||||||||||||:||||||||||||||||||||||||
Db        721 EWVVLLFLLLADARVCACLWMLILLGQAEAALEKLVVLHAASAASCNGFLYFVIFFVAAW 780

Qy        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
              ||||| |||| ||||||| | |||||||||||||||||| | |||::||||||||||||
Db        781 YIKGRAVPLAAYSLTGLWPFCLLLLALPQQAYAYDASVHGQTGVALLILITLFTLTPGYK 840

Qy        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
              |||| |||||||||||||:||| |||| |||||||||| :| ||||||||||||||:|
Db        841 ILLSRSLWWLCYLLTLGEAMIQEWVPPMQARGGRDGIIWAATMFCPGVVFDITKWLLAML 900

Qy        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
              ||||||: |||||||||||||||:||| ||||||||||||||||||||||||||:||
Db        901 GPAYLLRDALTRVPYFVRAHALLRLCTMVRHLAGGRYVQMALLALGRWTGTYIYDHLSPM 960

Qy        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
              ||||||||||||||||||||||||||||||||||||||||||:||||||||||||:||||
Db        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDVLHGLPVSARLGREILLGPAD 1020

Qy       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080
              ||||||| |||||||||||||||| ||||||||||||||||||||||||||||||||||
Db       1021 GYTSKGWRLLAPITAYAQQTRGLLGAIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080

Qy       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140

Qy       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200
              |||||||||||||||||||||||||||||||||||||||||:|||||||||||||||||
Db       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGIFRAAVCSRGVAKSIDFI 1200

Qy       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260
              ||| ||:|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 PVEALDVVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260

Qy       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGESITYSTYGKFLADGGCAGGAYDIIICD 1320

Qy       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380
              ||||||:|||||:|||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 ECHAVDATTILGVGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380

Qy       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440
              |||||||||||||||||||||||||:|||| ||||:|||||||||||||||||||||||
Db       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELSAALRSMGLNAVAYYRGLDVSVIPTQGDVVVVA 1440

Qy       1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db       1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQTVPQDAVSRSQRRGRTGRGRL 1500

Qy       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560

Qy       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
             |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLVAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620

Qy       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680
             |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TLVGPTPLLYRLGPVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680

Qy       1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740
             |||| |||||:||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TGCVSIIGRLHVNQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740

Qy       1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
             |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 QQASKQAQDIQPAVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800

Qy       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860

Qy       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920
             |||||||||||||||||||||:||||||||||||||||||||||||||||||||||||||||
Db       1861 AGISGALVAFKIMSGEKPSMEDVINLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920

Qy       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCAGSW 1980

Qy       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040
             ||:|||||||||||||||||||||||||:||||||:||||||||||||||||||||||||||
Db       1981 LREVWDWVCTILTDFKNWLTSKLFPKMPGLPFISCQKGYRGVWAGTGIMTTRCPCGANIS 2040

Qy       2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100
             |||||||||||||||| |||||||||||||||||||||| |||||||||||||||||||||
Db       2041 GNVRLGSMRITGPKTCMNTWQGTFPINCYTEGQCVPKPAPNFKTAIWRVAASEYAEVTQH 2100

Qy       2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160
             ||| ||||||||:||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 GSYSYITGLTTDSLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160

Qy       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220
             |||||||||| ||| |||||||||| ||||||||||||||||||||||||||||||||||||
Db       2161 GSQLPCDPEPDADVLTSMLTDPSHITAEAAARRLARGSPPSEASSSASQLSAPSLRATCT 2220

Qy       2221 THGKAYDVDMVDANLFMGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYMLPKKR 2280
             |||||||||||||||||||||||| |||||||||| |||||||||||||||||||||: |
Db       2221 THGKAYDVDMVDANLFMGGDVTRIESESKVVVLDSLDPMAEERSDLEPSIPSEYMLPRNR 2280

Qy       2281 FPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLSEDSI 2340
             |||||||||||||||||||||||||||| ||||||||||:||||||||||||||:|::|
Db       2281 FPPALPAWARPDYNPPLVESWKRPDYQPPTVAGCALPPPKKTPTPPPRRRRTVGLNENTI 2340

Qy       2341 GDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTPPDELALSETGSISSMPPLEGELG 2400
             ||||||||||:||||| |||||||||| |||||:||||| |||||||||||||||||| |
Db       2341 GDALQQLAIKAFGQPPLSGDSGLSTGADAADSGSRTPPDESALSETGSISSMPPLEGEPG 2400

Qy       2401 DPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460
             |||||||||| ||||||    ||||||||||||||||||||||||||||||||||||||||
Db       2401 DPDLEPEQVELQPPPQGEEVVPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460

Qy       2461 EKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIKLAASKV 2520
             |||||||||||||||||||||:|||||||||||||||||::|||||||||||||||||||||
Db       2461 EKLPINPLSNSLLRYHNKVYCTTSKSASLRAKKVTFDRMQVLDAHYDSVLKDIKLAASKV 2520

Qy       2521 TARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSETPIPTTI 2580
             :|||||:|||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db       2521 SARLLTLEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSQTPIPTTI 2580

Qy       2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPA 2640
             |||||||||| |||||||||||||||||||||||||||||:|||||||||||||||||||||
Db       2581 MAKNEVFCVDPAKGGKKAARLIVYPDLGVRVCEKMALYDVTQKLPQAVMGASYGFQYSPA 2640

Qy       2641 QRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTE 2700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2641 QRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTE 2700

Qy       2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPTMLVCGD 2760
             |||||||| |||||||||||||||||||||||||||:||||||||||:|||||||||||||
Db       2701 RLYVGGPMLNSKGQTCGYRRCRASGVLTTSMGNTITCYIKALAACKAAGIVAPTMLVCGD 2760

Qy       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVAXGPQG 2820

Qy       2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDT 2880
             |||||||||||||::|||||||||||||||||||||||||| ||||||||||||||||||
Db       2821 RRRYYLTRDPTTPLSRAAWETVRHSPVNSWLGNIIQYAPTIWVRMVLMTHFFSILMAQDT 2880

Qy       2881 LDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2940
             |||||||||||| |||||||||||||||||||||||||||:|||||||||||||||||||||
Db       2881 LDQNLNFEMYGXVYSVSPLDLPAIIERLHGLDAFSLHTYSPHELTRVASALRKLGAPPLR 2940

Qy       2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000
             ||||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||
Db       2941 AWKSRARAVRASLISRGGRAAICGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000

Qy       3001 DIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
             |||||||||||||||  |||| ||||||||||||
Db       3001 DIYHSVSRARPRLLLLSLLLLLVGVGLFLLPAR 3033
```

RESULT 4
Q99IB6
ID Q99IB6 PRELIMINARY; PRT; 3033 AA.
AC Q99IB6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21316767; PubMed=11424123; DOI=10.1002/jmv.1055;
RA Kato T., Furusaka A., Miyamoto M., Date T., Yasui K., Hiramoto J.,
RA Nagayama K., Tanaka T., Wakita T.;
RT "Sequence analysis of hepatitis C virus isolated from a fulminant
RT hepatitis patient.";
RL J. Med. Virol. 64:334-339(2001).
DR EMBL; AB047641; BAB32874.1; -.
DR HSSP; Q8JYS1; 1CWX.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR GO; GO:0019079; P:viral genome replication; IEA.
DR GO; GO:0019087; P:viral transformation; IEA.
DR InterPro; IPR000345; CytC_heme_BS.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR000745; HCV_NS4a.

```
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR004109; Peptidase_S29.
DR   InterPro; IPR009003; Pept_Ser_Cys.
DR   InterPro; IPR002518; Pept_U39_HCV_NS2.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; Transmembrane.
SQ   SEQUENCE   3033 AA;  328780 MW;  9AB16077CF46517B CRC64;

  Query Match             95.5%;  Score 15394;  DB 2;  Length 3033;
  Best Local Similarity   94.8%;  Pred. No. 0;
  Matches 2875;  Conservative   74;  Mismatches   84;  Indels    0;  Gaps    0;

Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSTTPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
              |||||||||| |:|:| |: ||| ||| |||||||| ||||||||||||||: |:|||
Db        181 LLSCITTPVSGAQVRNTSSSYMVANDCANDSITWQLDAAVLHVPGCVPCEKVGDTSRCWI 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
              ||||||||::|||||:||||||||||||||||||||||||||||||||||||||||| |||
Db        241 PVSPNVAVRQPGALTRGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQRHWF 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
              ||:|||||||| ::|||||||||||||||||||| ||||||||||||||||||||||||
Db        301 VQECNCSIYPGHVSGHRMAWDMMMNWSPTATMILAYVMRVPEVIIDIISGAHWGVMFGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
              |||||||||||||||||||| |: |||| | ||  : | |  | || |||:||||||
Db        361 YFSMQGAWAKVVVILLLAAGVDAGTYAVGGSTAHTTKGIASFFSPGARQNIQLINTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
              ||||||||||:||:|||:|:||| : |||||||||:||||:|| ||:||| ||||:||||
Db        421 HINRTALNCNESLNTGFLAALFYANRFNSSGCPERLSACRNIEDFRIGWGTLQYENNVTN 480

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
              |||||||:|||||||:| |::|||||||||||||||||| ||||||||||||||
Db        481 SEDMRPYCWYYPPRQCGIVPARSVCGPVYCFTPSPVVVGTTDRLGVPTYTWGENETDVFL 540

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
              ||||||| |||||||||:|:|||||||||| ||||||| ||||||||||||||  | :
Db        541 LNSTRPPSGSWFGCTWMNSTGFTKTCGAPPCRIRADFNASADLLCPTDCFRKHPDAAYFR 600

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
              ||||||||||||:|||||||||||||||||||||||||||||| ||||||||||:|||
Db        601 CGSGPWLTPRCLVDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLAAACNFTRGDRCDLED 660

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
              ||||||||||||||||||:|||||||||||||||||||||:||||||||||||||||
Db        661 RDRSQLSPLLHSTTEWAILPCTYSDLPALSTGLLHLHQNIVDVQYMYGLSPALTKYIVRW 720

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
              |||:||||||||||||:|||||||||||||||:||||||||||||||||||||||||||
Db        721 EWVVLLFLLLADARVCACVWMLILLGQAEAALEKLVVLHAASAASCNGFLYFVIFFVAAW 780

Qy        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
              ||||| |||||||||||||| |||||||||||||| | |||||| ||::|||||||| ||
Db        781 YIKGRAVPLATYSLTGLWSFCLLLLALPQQAYAYDTSAHGQIGATLLILITLFTLTPAYK 840

Qy        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
              ||:|||||||||||||||||||:|||||||||||| ||| |||||||||||:|
Db        841 TLVSRFLWWLCYLLTLGEAMVQEWAPPLQVRGGRDGIIWAAAIFCPGVVFDITKWLLALL 900

Qy        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
              ||||||: |||||||||||||||:||| ||||||||||||||||||||||||:||
Db        901 GPAYLLRDALTRVPYFVRAHALLRICTMVRHLAGGRYVQMALLALGRWTGTYIYDHLSPM 960

Qy        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
              |||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SDWAANGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020

Qy       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db       1021 GYTSKGWSLLAPITAYAQQTRGLLGAIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSIAG 1080

Qy       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVEWPSPPGTKSLEPCTCGAVDLYLVTRN 1140

Qy       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200
              |||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||
Db       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGIFRAAVCSRGVAKSIDFI 1200

Qy       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260

Qy       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320
              |||||||||||||||||||||||||||||| ||||||||||||:||||||||||
Db       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAAITYSTYGKFLADGGCSGGAYDIIICD 1320

Qy       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380
              |||:||:|||||||||||||||| ||||||||||||||||||||||||||||||||||
Db       1321 ECHSVDATTILGIGTVLDQAETAGVRSTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380

Qy       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440
              |||||||||||||||||||||||||||||||||||||||||:||||||||:||||||||||
Db       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNAVAYYRGLDVSIIPTQGDVVVVA 1440

Qy       1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500
              |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQTVPQDAVSRSQRRGRTGRGRL 1500

Qy       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560
              |||||||||||||||||||| |||||:|||||||:|||||||||||||||||||||||||
Db       1501 GIYRYVSTGERASGMFDSVVLSECYDSGAAWYELTPAETTVRLRAYFNTPGLPVCQDHLE 1560

Qy       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
              |||||||||||||||||||||||:|:|||||||||||||||||||||||||||||||||
Db       1561 FWEAVFTGLTHIDAHFLSQTKQAGDNFAYLTAYQATVCARAKAPPSSWDVMWKCLTRLKP 1620

Qy       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TLVGPTPLLYRLGPVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680

Qy       1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TGCVSIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740

Qy       1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
              |||||||||||| |||||||||||||||||| ||||||||||||||||||||||||||||
```

```
Db       1741 QQASKQAQDIQPAVQASWPKVEQFWAKHMWNFTSGIQYLAGLSTLPGNPAVASMMAFSAA 1800

Qy       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860

Qy       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920

Qy       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||| ||||| |||||||
Db       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEGCPIPCSGSW 1980

Qy       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040

Qy       2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100
             |||||||||||||||||| |||||||||||| |||||||||||| ||||||||||||||||||
Db       2041 GNVRLGSMRITGPKTCMNTWQGTFPINCYTGGQCVPKPAPNFKTAIWRVAASEYAEVTQH 2100

Qy       2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160
             ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 GSYSYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160

Qy       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220

Qy       2221 THGKAYDVDMVDANLFMGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYMLPKKR 2280
             |||| :|||||||||||||||||| ||||||||||||||  |||||||||||||||||||:
Db       2221 THGKTHDVDMVDANLFMGGDVTRIESESKVVVLDSLDPMVEEEGDLEPSIPSEYMLPKKK 2280

Qy       2281 FPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRRTVGLSEDSI 2340
             |||||| |||||||||||||  |:||||| ||||||||||:|||||||||||||||||| :|
Db       2281 FPPALPPWARPDYNPPLVESRKKPDYQPPTVAGCALPPPKKTPTPPPRRRRRTVGLSESTI 2340

Qy       2341 GDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTPPDELALSETGSISSMPPLEGELG 2400
             ||||||||||:|||||||||||||||| |||||||:||||| |||||||||||||||||| |
Db       2341 GDALQQLAIKAFGQPPPSGDSGLSTGADAADSGSRTPPDESALSETGSISSMPPLEGEPG 2400

Qy       2401 DPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460
             |||||||||| |||||||  |||||||||||||| |||||||||||||||||||||||||||
Db       2401 DPDLEPEQVERQPPPQGGEVAPGSDSGSWSTCSGEDDSVVCCSMSYSWTGALITPCSPEE 2460

Qy       2461 EKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIKLAASKV 2520
             ||||||||||||||||||||:||||||||||||||:||||||||||||||||||||||||||
Db       2461 EKLPINPLSNSLLRYHNKVYCTTSKSASLRAKKVTFDRMQVLDAYYDSVLKDIKLAASKV 2520

Qy       2521 TARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSETPIPTTI 2580
             ||:|||:|||||||||||||||||||||||||||||||||||||||||||:|||||||||||
Db       2521 TAKLLTLEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSQTPIPTTI 2580

Qy       2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPA 2640
             ||||||||||||||||||||||||||||||||||||||||||:||||||||||||||||||||
Db       2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDVTQKLPQAVMGASYGFQYSPA 2640

Qy       2641 QRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTE 2700
             ||||||||||||||||||||||:||||||||||||||||||||||||||||||||||||||||
Db       2641 QRVEFLLKAWAEKKDPMGFSYDTRCFDTTVTERDIRTEESIYRACSLPEEAHTAIHSLTE 2700

Qy       2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPTMLVCGD 2760
             ||||||||||||||:|||||||||||||||||||||| |||||||||||:|||||||||||
Db       2701 RLYVGGPMFNSKGQSCGYRRCRASGVLTTSMGNTITCYAKALAACKAAGIVAPTMLVCGD 2760

Qy       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820

Qy       2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDT 2880
             |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db       2821 RRRYYLTRDPTTPIARAAWETVRHSPANSWLGNIIQYAPTIWARMVLMTHFFSILMAQDT 2880

Qy       2881 LDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2940
             ||||||||||||||||||||||||||||||||||||||||||:||||||||||||||||||||
Db       2881 LDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYSPHELTRVASALRKLGAPPLR 2940

Qy       2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000

Qy       3001 DIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
             ||||||||||||||  |||| ||||||||||
Db       3001 DIYHSVSRARPRLLLLSLLLLSVGVGLFLLPAR 3033
```

```
RESULT 5
Q9IZA4
ID   Q9IZA4      PRELIMINARY;      PRT;  3033 AA.
AC   Q9IZA4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MD2a-4;
RA   Itakura J., Nagayama K., Enomoto N., Kurosaki M., Watanabe H.,
RA   Sato C.;
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF238483; AAF59942.1; -.
DR   HSSP; Q8JYS1; 1CWX.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR011545; DEAD/DEAH_N.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_S29.
DR   InterPro; IPR009003; Pept_Ser_Cys.
DR   InterPro; IPR002518; Pept_U39_HCV_NS2.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
```

```
DR   Pfam; PF00271; Helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; Transmembrane.
SQ   SEQUENCE   3033 AA;  329421 MW;  5EAAECD220F4D5CB CRC64;

  Query Match           95.3%;  Score 15356;  DB 2;  Length 3033;
  Best Local Similarity   94.4%;  Pred. No. 0;
  Matches 2864;  Conservative   81;  Mismatches   88;  Indels    0;  Gaps    0;

Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPTDPRHRSRNVG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
              |||||||||||||||||||||||||||||| |||||||||||||:|||||||||||||||
Db        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARTLAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
              |||||| |||| |||| | ||||||||:| ||||||||||||||||||||||| |:|||
Db        181 LLSCITIPVSAIEVKNTSDGYMVTNDCSNGSITWQLQAAVLHVPGCVPCEKVGNTSRCWI 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||:||||||
Db        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAHMFIVTPQHHWF 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIIGGAHWGVMFGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
              |||||||||||||||| ||||| |::|||| |  |      || :| :| |||:||||||
Db        361 YFSMQGAWAKVVVILLLTAGVDANTYSVGGSTAHLTQTFAGLFTLGSQQNIQLINTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
              ||||||||||:||:|||:|||||||:|||||||||:||||:|||||:||| ||||:||||
Db        421 HINRTALNCNESLNTGFLASLFYTHAFNSSGCPERLSACRNIEAFRIGWGVLQYEENVTN 480

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
              |||||||||||||:|||:| |::|||||||||||||||||:|| ||||||||||||||||
Db        481 PEDMRPYCWHYPPKQCGIVPARSVCGPVYCFTPSPVVVGTTDKLGVPTYTWGENETDVFL 540

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
              ||||||| |:|||||||:|:|||||||||| |||||||||||||||||||: || | |
Db        541 LNSTRPPRGAWFGCTWMNSTGFTKTCGAPPCRIRADFNASTDLLCPTDCFRKHPEATYTK 600

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
              ||||||||||::|||||| :|||||||||||||||||||||||||||||||||||||||
Db        601 CGSGPWLTPRCMVDYPYRLGIHPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
              ||||||||||||||||||::|||||||||||||||||||:|||||||:|::|||||||||
Db        661 RDRSQLSPLLHSTTEWAILPCTFSDLPALSTGLLHLHQNIVDVQYMYGLSPAITRHIVRW 720

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
              |||:|||||||||||:|||||||||||||:|||||||||||||||||||||||||||||
Db        721 EWVVLLFLLLADARVCACVWMLILLGQAEAALEKLVVLHAASAASCNGFLYFVIFFVAAW 780

Qy        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
              ||||| |||| |||||||| |||||||||||||||||||| ||:::|||||||||||||
Db        781 YIKGRAVPLAAYSLTGLWSFCLLLLALPQQAYAYDASVHGQIGAVLLILVTLFTLTPGYK 840

Qy        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
              |||| |||||||||||||:||| |||||||||||||  || ||||||||||||:|
Db        841 ILLSRCLWWLCYLLTLGEAMIQEWVPPMQVRGGRDGIIWAATIFCPGVVFDITKWLLALL 900

Qy        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
              || |||:|||||||||||||||||||||||||:|||||||||||||||||||||||||||
Db        901 GPGYLLRGALTRVPYFVRAHALLRMCTMARHLAGGKYVQMALLALGRWTGTYIYDHLTPM 960

Qy        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020

Qy       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080
              ||||||| |||||||||||||||| ||||||||||||||||||||||||||||||||||
Db       1021 GYTSKGWKLLAPITAYAQQTRGLLGAIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080

Qy       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140

Qy       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200
              |||||||||||||||||||||||||||||||||||||||||:|||||||||||||||||
Db       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGIFRAAVCSRGVAKSIDFI 1200

Qy       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Db       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLILNP 1260

Qy       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320
              || |||||||||||:|||||||||||||||||||||||||||||||||||||||||||||
Db       1261 SVGATLGFGAYLSKAYGINPNIRTGVRTVTTGDSITYSTYGKFLADGGCAGGAYDIIICD 1320

Qy       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380
              ||||| ||| ||||||||||||||||||||| |||| ||||||||||||||||||||||
Db       1321 ECHAVGSTTNLGIGTVLDQAETAGVRLTVLATATLPGSVETPHPNIEEVALGQEGEIPFY 1380

Qy       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440
              |||||||||||||||||||||||||||||||||||||||:|||||||||||||||||||||
Db       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNAVAYYRGLDVSVIPTQGDVVVVA 1440

Qy       1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500
              |||||||:|||||:|||||||||||||||||||||||||| |||||||||||||||||||
Db       1441 TDALMTGFTGDFDTVIDCNVAVTQVVDFSLDPTFTITTQTVPQDAVSRSQRRGRTGRGRL 1500

Qy       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560
              | ||||||||||||||||||||||||||||||||:|||||||||||||||||||||||||
Db       1501 GTYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPAETTVRLRAYFNTPGLPVCQDHLE 1560

Qy       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
              |||||||||||||||||||:||:| || |||||||||||||||:||||||||||||||||
Db       1561 FWEAVFTGLTHIDAHFLSQTKQAGESFPYLVAYQATVCARAKAPPPSWDIMWKCLTRLKP 1620

Qy       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680
              ||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||
Db       1621 TLVGPTPLLYRLGSVTNEITLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680

Qy       1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740
              |||| |||||||||||:|||||||||||||||||||:|||:|||||||||||||||||||
Db       1681 TGCVSIIGRLHINQRAIVAPDKEVLYEAFDEMEECASKAALLEEGQRIAEMLKSKIQGLL 1740

Qy       1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
              |||||||||||| |||||||||||||||||||||||||| |||||||||||||||||||
Db       1741 QQASKQAQDIQPAVQASWPKVEQFWAKHMWNFISGIQYLARLSTLPGNPAVASMMAFSAA 1800

Qy       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860
              |||||||||||| |||||||||||||||||||||||||||||||||:|||||||||||||
Db       1801 LTSPLSTSTTILLNILGRWLASQIAPPAGATGFVVSGLVGAAVGNIGLGKVLVDILAGYG 1860

Qy       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920
              |||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||
Db       1861 AGISGALVAFKIMSGEKPSMEDVINLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920

Qy       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCAGSW 1980

Qy       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040
              |||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||
```

```
Db       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFISCQKGYKGVWAGTGIMTTRCPCGANIS 2040

Qy       2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db       2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKTAIWRVAASEYAEVTQH 2100

Qy       2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160
              | | |:|||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db       2101 GQYSYVTGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFYVGLNSFVV 2160

Qy       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220
              ||||||||||||||||||:||||||| ||||||||||||||||||||||||||||||||||
Db       2161 GSQLPCDPEPDTDVLMSMLTEPSHITAEAAARRLARGSPPSEASSSASQLSAPSLRATCT 2220

Qy       2221 THGKAYDVDMVDANLFMGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYMLPKKR 2280
              |||||||||||||||||||||||| |||||||||| ||:|||||||||||||||: |
Db       2221 THGKAYDVDMVDANLFMGGDVTRIESESKVVVLDSLDPMAEEKSDLEPSIPSEYMLPRNR 2280

Qy       2281 FPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLSEDSI 2340
              |||||||||||||||||||||||| |||||||||:|||||||||||||| | :|
Db       2281 FPPALPAWARPDYNPPLVESWKRPDYQPPTVAGCALPPPKKTPTPPPRRRRTVGLDESTI 2340

Qy       2341 GDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTPPDELALSETGSISSMPPLEGELG 2400
              ||||||||||:||| ||| |||||||| ||||| :| |||||||||||||||||||| |
Db       2341 GDALQQLAIKTFGQIPPSADSGLSTGADAADSGGRTSPDELALSETGSISSMPPLEGEPG 2400

Qy       2401 DPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460
              |||||||||| ||||| |  ||||||||||||||||||||||||||||||||||||||||
Db       2401 DPDLEPEQVELQPPPQEGELAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460

Qy       2461 EKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIKLAASKV 2520
              |||||||||||||||||||||:|||||||||||||||||::|||||||||||||||||||
Db       2461 EKLPINPLSNSLLRYHNKVYCTTSKSASLRAKKVTFDRMQVLDAHYDSVLKDIKLAASKV 2520

Qy       2521 TARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSETPIPTTI 2580
              :|||||:|||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db       2521 SARLLTLEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSQTPIPTTI 2580

Qy       2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPA 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPA 2640

Qy       2641 QRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTE 2700
              |||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||
Db       2641 QRVEFLLKAWAEKKDPLGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTE 2700

Qy       2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPTMLVCGD 2760
              |||||||| |||||:|||||||||||||||||||||||||||||||||:|||||||||
Db       2701 RLYVGGPMLNSKGQSCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIVAPTMLVCGD 2760

Qy       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| | |
Db       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALNPLG 2820

Qy       2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDT 2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDT 2880

Qy       2881 LDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2940
              ||||||||||:|||||||||||||||||||||||||||:||||||||| ||||||||||
Db       2881 LDQNLNFEMYGSVYSVSPLDLPAIIERLHGLDAFSLHTYSPHELTRVASTLRKLGAPPLR 2940

Qy       2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000

Qy       3001 DIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
              ||||||||||||||| ||||| ||||||||||
Db       3001 DIYHSVSRARPRLLLLGLLLLSVGVGLFLLPAR 3033

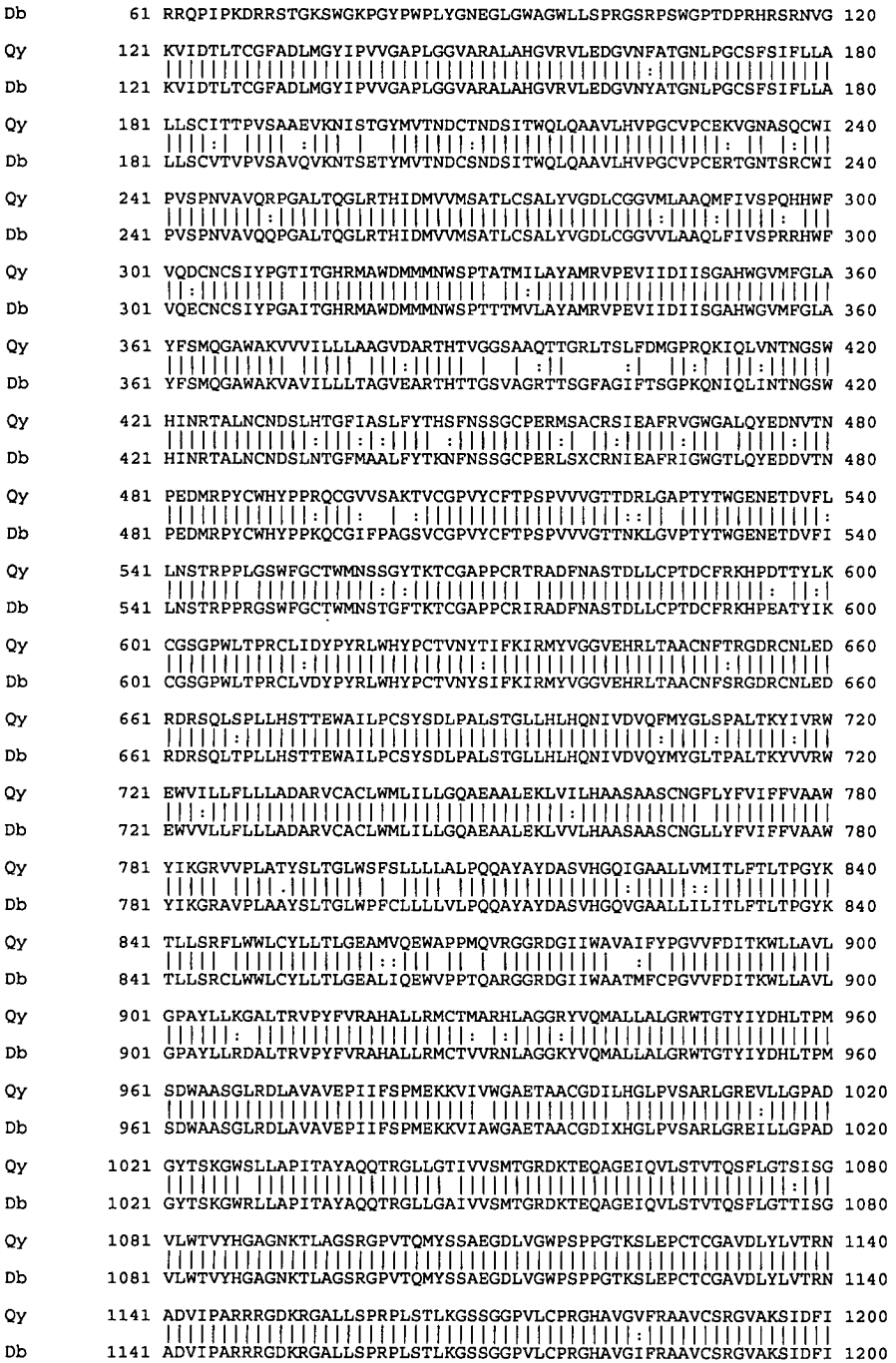

RESULT 6
Q9IZA5
```

```
ID   Q9IZA5      PRELIMINARY;      PRT;  3033 AA.
AC   Q9IZA5;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MD2a-2;
RA   Itakura J., Nagayama K., Enomoto N., Kurosaki M., Watanabe H.,
RA   Sato C.;
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF238482; AAF59941.1; -.
DR   HSSP; Q8JYS1; 1CWX.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR011545; DEAD/DEAH_N.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR004109; Peptidase_S29.
DR   InterPro; IPR009003; Pept_Ser_Cys.
DR   InterPro; IPR002518; Pept_U39_HCV_NS2.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; Transmembrane.
SQ   SEQUENCE   3033 AA;  329099 MW;  74E61003AEC7816A CRC64;

  Query Match             95.3%;  Score 15355;  DB 2;  Length 3033;
  Best Local Similarity   94.2%;  Pred. No. 0;
  Matches 2857;  Conservative   83;  Mismatches   93;  Indels    0;  Gaps    0;

Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
```


Db 61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPTDPRHRSRNVG 120
Qy 121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
Db 121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180
Qy 181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
Db 181 LLSCVTVPVSAVQVKNTSETYMVTNDCSNDSITWQLQAAVLHVPGCVPCERTGNTSRCWI 240
Qy 241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
Db 241 PVSPNVAVQQPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVVLAAQLFIVSPRRHWF 300
Qy 301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
Db 301 VQECNCSIYPGAITGHRMAWDMMMNWSPTTTMVLAYAMRVPEVIIDIISGAHWGVMFGLA 360
Qy 361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
Db 361 YFSMQGAWAKVAVILLLTAGVEARTHTTGSVAGRTTSGFAGIFTSGPKQNIQLINTNGSW 420
Qy 421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
Db 421 HINRTALNCNDSLNTGFMAALFYTKNFNSSGCPERLSXCRNIEAFRIGWGTLQYEDDVTN 480
Qy 481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
Db 481 PEDMRPYCWHYPPKQCGIFPAGSVCGPVYCFTPSPVVVGTTNKLGVPTYTWGENETDVFI 540
Qy 541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
Db 541 LNSTRPPRGSWFGCTWMNSTGFTKTCGAPPCRIRADFNASTDLLCPTDCFRKHPEATYIK 600
Qy 601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
Db 601 CGSGPWLTPRCLVDYPYRLWHYPCTVNYSIFKIRMYVGGVEHRLTAACNFSRGDRCNLED 660
Qy 661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
Db 661 RDRSQLTPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQYMYGLTPALTKYVVRW 720
Qy 721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
Db 721 EWVVLLFLLLADARVCACLWMLILLGQAEAALEKLVVLHAASAASCNGLLYFVIFFVAAW 780
Qy 781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
Db 781 YIKGRAVPLAAYSLTGLWPFCLLLLVLPQQAYAYDASVHGQVGAALLILITLFTLTPGYK 840
Qy 841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
Db 841 TLLSRCLWWLCYLLTLGEALIQEWVPPTQARGGRDGIIWAATMFCPGVVFDITKWLLAVL 900
Qy 901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
Db 901 GPAYLLRDALTRVPYFVRAHALLRMCTVVRNLAGGKYVQMALLALGRWTGTYIYDHLTPM 960
Qy 961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
Db 961 SDWAASGLRDLAVAVEPIIFSPMEKKVIAWGAETAACGDIXHGLPVSARLGREILLGPAD 1020
Qy 1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080
Db 1021 GYTSKGWRLLAPITAYAQQTRGLLGAIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTTISG 1080
Qy 1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
Db 1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
Qy 1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200
Db 1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGIFRAAVCSRGVAKSIDFI 1200
Qy 1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260
Db 1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260
Qy 1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320
Db 1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAAITYSTYGKFLADGGCAGGAYDIIICD 1320
Qy 1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380
Db 1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380
Qy 1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440
Db 1381 GRAIPLSHIKGGRHLIFCHSKKKCDELAAALRGMGLNAVAYYRGLDVSVIPTQGDVVVFA 1440
Qy 1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500
Db 1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQTVPQDAVSRSQRRGRTGRGRL 1500
Qy 1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560
Db 1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPAETTVRLRAYFNTPGLPVCQDHLE 1560
Qy 1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
Db 1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLVAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
Qy 1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680
Db 1621 TLVGPTPLLYRLGSVANEITLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680
Qy 1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740
Db 1681 TGCVSIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALVEEGQRIAEMLKSKIQGLL 1740
Qy 1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
Db 1741 QQASKQAQDIQPAVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
Qy 1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860
Db 1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860
Qy 1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920
Db 1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920
Qy 1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980
Db 1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCAGSW 1980
Qy 1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040
Db 1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFISCQKGYKGVWAGTGIMTTRCPCGANIS 2040
Qy 2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100
Db 2041 GNVRLGSMRITGPKTCMNTWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100
Qy 2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160
Db 2101 GSYSYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160
Qy 2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220
Db 2161 GSQLPCDPEPDTDVLTSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220
Qy 2221 THGKAYDVDMVDANLFMGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYMLPKKR 2280
Db 2221 THGKTYDVDMVDASLFMGGDVTRIESESKVVVLDSLDPMAEEESDLEPSIPSEYMLSRNR 2280

```
Qy       2281 FPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLSEDSI 2340
             ||||||||||||||||||||||||||||||||||||||||||||:| ||||||||||||||| :|
Db       2281 FPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPKKAPTPPPRRRRTVGLSESTI 2340

Qy       2341 GDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTPPDELALSETGSISSMPPLEGELG 2400
             |||||||||||||||||||||||| ||||||::|||||||||||| ||||||||| |
Db       2341 GDALQQLAIKSFGQPPPSGDSGLSTGADAADSGSRSPPDELALSETGSTSSMPPLEGEPG 2400

Qy       2401 DPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460
             |||||||||||||||||  ||||||||||||||: ||| ||:|||||| ||||| ||:
Db       2401 DPDLEPEQVEPQPPPQGGEVAPGSDSGSWSTCSEENVSVVFCSISYSWTGGLITPCIPEK 2460

Qy       2461 EKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIKLAASKV 2520
             |||| |||||||||| |||||||:|||||||||||||||||:|||||||||||||||||||
Db       2461 EKLPFNPLSNSLLRYDNKVYCTTSKSASLRAKKVTFDRMQVLDAYYDSVLKDIKLAASKV 2520

Qy       2521 TARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSETPIPTTI 2580
             :|||||:|||||||||||||||||||||||||||||||||||||||||||:|||||||
Db       2521 SARLLTLEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSQTPIPTTI 2580

Qy       2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPA 2640
             |||||||||| |||||||||||||||||||||||||:|||||||||||||||||||||||
Db       2581 MAKNEVFCVDPNKGGKKAARLIVYPDLGVRVCEKMALYDVTQKLPQAVMGASYGFQYSPA 2640

Qy       2641 QRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTE 2700
             |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db       2641 QRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRPCSLPEEAHTAIHSLTE 2700

Qy       2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPTMLVCGD 2760
             |||||||||||||| ||||||||||||||||||||||||||||||||:|||||||||
Db       2701 RLYVGGPMFNSKGQACGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIVAPTMLVCGD 2760

Qy       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820
             |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVHLGPQG 2820

Qy       2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDT 2880
             |||||||||||||||||||||||||||||||||||||||:| ||||||||||||||||||
Db       2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTVWVRMVLMTHFFSILMAQDT 2880

Qy       2881 LDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2940
             ||||||||||:|||||||||||||||||||||||||:||||||||:||||||||||||||
Db       2881 LDQNLNFEMYGSVYSVSPLDLPAIIERLHGLDAFSLHTYSPHELTRVAAALRKLGAPPLR 2940

Qy       2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000
             |||||||||||||| |||||||||||||||||||||||||||||| ||||||||||||||
Db       2941 AWKSRARAVRASLISHGGRAAVCGRYLFNWAVKTKLKLTPLPEARHLDLSSWFTVGAGGG 3000

Qy       3001 DIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
              |||||||||||||| ||||||||||||||||
Db       3001 HIYHSVSRARPRLLLLGLLLLFVGVGLFLLPAR 3033


RESULT 7
Q99IB7
ID   Q99IB7      PRELIMINARY;      PRT;  3033 AA.
AC   Q99IB7;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21316767; PubMed=11424123; DOI=10.1002/jmv.1055;
RA   Kato T., Furusaka A., Miyamoto M., Date T., Yasui K., Hiramoto J.,
RA   Nagayama K., Tanaka T., Wakita T.;
RT   "Sequence analysis of hepatitis C virus isolated from a fulminant
RT   hepatitis patient.";
RL   J. Med. Virol. 64:334-339(2001).
DR   EMBL; AB047640; BAB32873.1; -.
DR   HSSP; Q8JYS1; 1CWX.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR011545; DEAD/DEAH_N.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR004109; Peptidase_S29.
DR   InterPro; IPR009003; Pept_Ser_Cys.
DR   InterPro; IPR002518; Pept_U39_HCV_NS2.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; Transmembrane.
SQ   SEQUENCE   3033 AA;  329035 MW;  74BD3B786F7D77D8 CRC64;

  Query Match             95.2%;  Score 15354;  DB 2;  Length 3033;
  Best Local Similarity   94.4%;  Pred. No. 0;
  Matches 2863;  Conservative   73;  Mismatches   97;  Indels    0;  Gaps    0;

Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKASERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RRQPIPKHRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
              |||||||||||||:||:||||||| ||| ||||||||||||||||||||||||||||||
Db        121 KVIDTLTCGFADLLGYVPVVGAPLSGVASALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
              |||||||||||| :||| |  || ||||:||||||:||||||||||||:|| |:|||
Db        181 LLSCITTPVSAVQVKNTSNAYMATNDCSNDSITWQLEAAVLHVPGCVPCEKMGNTSRCWI 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
              ||||||||::||||:||||||||||:||||||||||||||||||||:||||||||||||
Db        241 PVSPNVAVRQPGALTRGLRTHIDMVVLSATLCSALYVGDLCGGVMLASQMFIVSPQHHWF 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
              ||:|||||||| |||||||||||||||| |||||| ||||||||||||||||||||||||
```

```
Db        301 VQECNCSIYPGAITGHRMAWDMMMNWSPTTTMILAYVMRVPEVIIDIISGAHWGVMFGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
              ||||||||||||||||||:|||| | | | :| :||  | | |  | || |||:||||||
Db        361 YFSMQGAWAKVVVILLLASGVDAYTTTTGSAAGRTTSSLASAFSPGARQNIQLINTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
              |||||||||||||||||  :||| | |||||||||:||||:|| ||:|||||||:|||||
Db        421 HINRTALNCNDSLHTGFFTALFYIHKFNSSGCPERLSACRNIEDFRIGWGALQYDDNVTN 480

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
              |||||||||||||:||||| | ||||||||||||||||||||| ||||||||||||||||
Db        481 PEDMRPYCWHYPPKQCGVVPAGTVCGPVYCFTPSPVVVGTTDRLGVPTYTWGENETDVFL 540

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
              ||||||| |||||||||||:|:|||||||||||||||| |||||||||||||: ||:|
Db        541 LNSTRPPSGSWFGCTWMNSTGFTKTCGAPPCRTRADFNTSTDLLCPTDCFRKHPEATYIK 600

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
              |||||||||:||:|||||||||||||||: |||||||||||||| ||||||||||||||||
Db        601 CGSGPWLTPKCLVDYPYRLWHYPCTVNYSTFKIRMYVGGVEHRLMAACNFTRGDRCNLED 660

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
              ||||| :|||||||||||||||:|||||||||||||||||||:|||||||||:|||||
Db        661 RDRSQQTPLLHSTTEWAILPCSFSDLPALSTGLLHLHQNIVDVQYMYGLSPALTQYIVRW 720

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
              |||:|||||||||||||||||||||||||||||||:|||||||||||||||||| ||||
Db        721 EWVVLLFLLLADARVCACLWMLILLGQAEAALEKLVVLHAASAASCNGFLYFVIFLVAAW 780

Qy        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
              :||||||||| ||||||| | |||||||||||||||||:|||||||:||||||||||||
Db        781 HIKGRVVPLAAYSLTGLWPFCLLLLALPQQAYAYDASVHGQVGAALLVLITLFTLTPGYK 840

Qy        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
              ||||: |||||||||| | ||||||| || |||||||||  || ||||||||||||||||
Db        841 TLLSQSLWWLCYLLTLAETMVQEWAPSMQARGGRDGIIWAATIFCPGVVFDITKWLLAVL 900

Qy        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
              || |||:|||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        901 GPGYLLRGALTRVPYFVRAHALLRMCTMVRHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960

Qy        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
              |||||||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREILLGPAD 1020

Qy       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080
              ||||||| ||||||||||||||||:||||||||||||||:||||||||||||||||||
Db       1021 GYTSKGWKLLAPITAYAQQTRGLLGSIVVSMTGRDKTEQAGEVQVLSTVTQSFLGTSISG 1080

Qy       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140

Qy       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200
              |||||||||||||||||||||||||||||||||||||||||:||||||||||||||||||
Db       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGIFRAAVCSRGVAKSIDFI 1200

Qy       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260

Qy       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGEPITYSTYGKFLADGGCAGGAYDIIICD 1320

Qy       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380
              |||:||:|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 ECHSVDATTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380

Qy       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440
              ||| |||||||||||||||||||||| ||||||||:||||||||:||||||||||||||
Db       1381 GRAFPLSYIKGGRHLIFCHSKKKCDELATALRGMGLNAVAYYRGLDVSIIPTQGDVVVVA 1440

Qy       1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500
              |||||||||||||||||||||||||| |||||||||||||| ||||||||||||||||||||
Db       1441 TDALMTGYTGDFDSVIDCNVAVTQAVDFSLDPTFTITTQTVPQDAVSRSQRRGRTGRGRL 1500

Qy       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560
              ||||||||||||||||||||||||||||||||||:| ||||||||||||||||||||||
Db       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELSPVETTVRLRAYFNTPGLPVCQDHLE 1560

Qy       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLVAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620

Qy       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680

Qy       1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740
              |||| |||||||||||||||||||||||||||||||||||||:|||||||||||||||||
Db       1681 TGCVSIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALLEEGQRIAEMLKSKIQGLL 1740

Qy       1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
              |||||||||||| |||||||:|||||||||||||||||||||||||||||||||||||||
Db       1741 QQASKQAQDIQPAVQASWPKMEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800

Qy       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860

Qy       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920
              |||||||||||||||||||||||:|||||||||||||||||||||||||||||||||||
Db       1861 AGISGALVAFKIMSGEKPSMEDVINLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920

Qy       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCAGSW 1980

Qy       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040
              |||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||
Db       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFISCQKGYKGVWAGTGIMTTRCPCGANIS 2040

Qy       2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100
              |||||||||||||||||||||||||||||||:|||| ||| |||||||||||||||||||
Db       2041 GNVRLGSMRITGPKTCMNTWQGTFPINCYTEGQCLPKPALNFKTAIWRVAASEYAEVTQH 2100

Qy       2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160
              ||| |||||||||||||||||||||||||||||| |||||||||||||| ||||||||
Db       2101 GSYAYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRSAPTPKPFFRDEVSFSVGLNSFVV 2160

Qy       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220
              |||||||||||:|:|||||||||||| ||||||||||||||||||||||||||||||||
Db       2161 GSQLPCDPEPDTEVVMSMLTDPSHITAEAAARRLARGSPPSEASSSASQLSAPSLRATCT 2220

Qy       2221 THGKAYDVDMVDANLFMGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYMLPKKR 2280
              |||: |||||||||||||| | |||| |||||||||| | ||  ||||||:|||||:||
Db       2221 THGRTYDVDMVDANLFMGGGVIRIESESKVVVLDSLDSMTEEEGDLEPSVPSEYMLPRKR 2280

Qy       2281 FPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLSEDSI 2340
              |||||||||||||||||||||||||||| ||||||||:|||||||||||||||| :|
Db       2281 FPPALPAWARPDYNPPLVESWKRPDYQPPTVAGCALPPPKKTPTPPPRRRRTVGLSESTI 2340

Qy       2341 GDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTPPDELALSETGSISSMPPLEGELG 2400
              |||||||||||||||||||||||||| ||||| :|||||||||||| ||||||||| |
Db       2341 GDALQQLAIKSFGQPPPSGDSGLSTGADAADSGDRTPPDELALSETGSTSSMPPLEGEPG 2400

Qy       2401 DPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460
              |||||||||| ||||||| ||||||||||||||||||||||||||||||||||||||||
Db       2401 DPDLEPEQVELQPPPQGGEAAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460

Qy       2461 EKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIKLAASKV 2520
              |||||| |||||||||||||:|||||||||||||||||:|||||||||||||||||||||
Db       2461 EKLPINSLSNSLLRYHNKVYCTTSKSASLRAKKVTFDRMQVLDAYYDSVLKDIKLAASKV 2520
```

```
Qy       2521 TARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSETPIPTTI 2580
              :|||||:|||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db       2521 SARLLTLEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSQTPIPTTI 2580

Qy       2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPA 2640
              |||||||||| ||||| |||||||||||||||||||:||||||||||||||||||||||
Db       2581 MAKNEVFCVDPAKGGKKPARLIVYPDLGVRVCEKMALYDVTQKLPQAVMGASYGFQYSPA 2640

Qy       2641 QRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTE 2700
              ||||||||||||:|||||||||||||||||||||||||||:|||||||| ||||||||
Db       2641 QRVEFLLKAWAEKRDPMGFSYDTRCFDSTVTERDIRTEESIYQACSLPEEARTAIHSLTE 2700

Qy       2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPTMLVCGD 2760
              |||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||
Db       2701 RLYVGGPMFNSKGQSCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPTMLVCGD 2760

Qy       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820

Qy       2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDT 2880
              |||||||||||| |||||||||||||||||||||||||| ||||||||||||||||||
Db       2821 RRRYYLTRDPTTSIARAAWETVRHSPVNSWLGNIIQYAPTIWVRMVLMTHFFSILMAQDT 2880

Qy       2881 LDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2940
              ||||||||||:|||||||||||||||||||||||||||||||||||||||||||||||||
Db       2881 LDQNLNFEMYGSVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2940

Qy       2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000

Qy       3001 DIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
              |||||||||||||||  |||| ||||||||||
Db       3001 DIYHSVSRARPRLLLLSLLLLSVGVGLFLLPAR 3033
```

```
RESULT 8
Q99IB3
ID   Q99IB3      PRELIMINARY;      PRT;  3033 AA.
AC   Q99IB3;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21316767; PubMed=11424123; DOI=10.1002/jmv.1055;
RA   Kato T., Furusaka A., Miyamoto M., Date T., Yasui K., Hiramoto J.,
RA   Nagayama K., Tanaka T., Wakita T.;
RT   "Sequence analysis of hepatitis C virus isolated from a fulminant
RT   hepatitis patient.";
RL   J. Med. Virol. 64:334-339(2001).
DR   EMBL; AB047644; BAB32877.1; -.
DR   HSSP; Q8JYS1; 1CWX.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR011545; DEAD/DEAH_N.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR004109; Peptidase_S29.
DR   InterPro; IPR009003; Pept_Ser_Cys.
DR   InterPro; IPR002518; Pept_U39_HCV_NS2.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; Transmembrane.
SQ   SEQUENCE   3033 AA;  328908 MW;  B306821483E70CB3 CRC64;

  Query Match             95.2%;  Score 15350;  DB 2;  Length 3033;
  Best Local Similarity   94.3%;  Pred. No. 0;
  Matches 2860;  Conservative   81;  Mismatches   92;  Indels    0;  Gaps    0;

Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              |||||||||||||||||||||||||:|||||||||||||||||||||||||||||||||
Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGRIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPTDPRHRSRNVG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
              |||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KVIDTITCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
              |||||| |||| ||:||| ||||||:|:||||||:|||||||||||||| |:|||
Db        181 LLSCITIPVSAVEVRNISDSYMVTNDCSNNSITWQLEAAVLHVPGCVPCEKVGNTSRCWI 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
              ||||||||::|||||||| |||||||||||||:|||||||||||| ||||||||||
Db        241 PVSPNVAVRQPGALTQGLRAHIDMVVMSATLCSALHVGDLCGGVMLAAQTFIVSPQHHWF 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
              ||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 VQECNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
              |||||||||||||||||||| ||| |  ||     | |||  ||:| |||:||||||
Db        361 YFSMQGAWAKVVVILLLAAGVDAGTHTTGAVAASNARGLASLFTSGPKQNIQLINTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
              |||||||||||| ||||::| | : |||||||||:| ||:|||||:|||||||||||
Db        421 HINRTALNCNDSLQTGFISALIYRNRFNSSGCPERLSVCRNIEAFRIGWGALQYEDNVTN 480

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
              |||||||||||| |||:| |::|||||||||||||||||| ||||||||||:|:|||
Db        481 PEDMRPYCWHYPPNQCGIVPAQSVCGPVYCFTPSPVVVGTTDRRGAPTYTWGENDTNVFL 540

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
              |||||||||||||||||:|:|||||||||| ||||| |||||||||||||| ||:|
```

```
Db        541 LNSTRPPLGSWFGCTWMNSTGFTKTCGAPPCRIRADFNTSTDLLCPTDCFRKHPDATYIK 600

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
              ||||||||||:||:||||||||||||||||||:||||||||||||||||||||||||||||
Db        601 CGSGPWLTPKCLVDYPYRLWHYPCTVNYTIFKVRMYVGGVEHRLTAACNFTRGDRCNLED 660

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
              |||||  ||||||||||:||||:||||||||||||||||||:||||||||||:|||
Db        661 RDRSQQSPLLHSTTEWAVLPCSFSDLPALSTGLLHLHQNIVDVQYMYGLSPALTKYVVRW 720

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
              |||:||||||||||||:||||||||||||||:|||||||||| ||||||||||||
Db        721 EWVVLLFLLLADARVCACVWMLILLGQAEAALEKLVVLHAASAASCNSFLYFVIFFVAAW 780

Qy        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
              |||||  ||||  |||||||  | ||||||||||||||||||||||::||||||||||
Db        781 YIKGRAVPLAAYSLTGLWPFCLLLLALPQQAYAYDASVHGQIGAALLILITLFTLTPGYK 840

Qy        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
              |||||  |||||||||||||||||||||||  ||||||||||  |  ||||||||||||||
Db        841 TLLSRSLWWLCYLLTLGEAMVQEWAPPMQARGGRDGIIWAATILCPGVVFDITKWLLAVL 900

Qy        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
              ||||||: ||| |||||||||||||||| :|||||||||||||||:||||||||||||
Db        901 GPAYLLRDALTSVPYFVRAHALLRMCTMVKHLAGGRYVQMALLALGRWSGTYIYDHLTPM 960

Qy        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
              ||||||||||||||||||||||||||||||||||||||||||:||||||||||:|||| |
Db        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILYGLPVSARLGREILLGPTD 1020

Qy       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080
              |||||||  |||||||||||||||| ||||||||||||||||||||:||||||||||
Db       1021 GYTSKGWRLLAPITAYAQQTRGLLGAIVVSMTGRDKTEQAGEIQVLSTVSQSFLGTSISG 1080

Qy       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
              ||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 VLWTVYHGTGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140

Qy       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200
              ||||||||||||||||||||||||||||||||||||||||||:||||||||||||||||||
Db       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGIFRAAVCSRGVAKSIDFI 1200

Qy       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260
              |||  |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 PVEALDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260

Qy       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320
              |||||||||||||||||||||||||||||||||  ||||||||||||||:||||||||||
Db       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGGPITYSTYGKFLADGGCSGGAYDIIICD 1320

Qy       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQGEGEIPFY 1380
              ||||||:||||||||||||||||||||||||||||||||||||||:||||||||||||||
Db       1321 ECHAVDATTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEIALGQGEGEIPFY 1380

Qy       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440
              |||||||:|||||||||||||||||||||||||||||||:||||||||||:||||||||||
Db       1381 GRAIPLSHIKGGRHLIFCHSKKKCDELAAALRGMGLNAVAYYRGLDVSIIPTQGDVVVVA 1440

Qy       1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500
              |||||||: |||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       1441 TDALMTGFAGDFDSVIDCNVAVTQVVDFSLDPTFTITTQTVPQDAVSRSQRRGRTGRGRL 1500

Qy       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560
              | |||||||||||||||||||||||||||||||||||||| ||||||  ||||||||||||||||
Db       1501 GTYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPVETTVRLGAYFNTPGLPVCQDHLE 1560

Qy       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLVAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620

Qy       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYWLA 1680

Qy       1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740
              ||||  |||||||||||||||||||||||||||||||||||||||||||||:||||||||||
Db       1681 TGCVSIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRMAEMLKSKIQGLL 1740

Qy       1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
              ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 QQASKQAQDIQPAVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800

Qy       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860
              |||||||||||||||||||||||||||||||||||||||||||||||:||||||||||||
Db       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSVGLGKVLVDILAGYG 1860

Qy       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 TGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920

Qy       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980
              ||||||||  |||||||||||||||||||||||||||||||||||||||||||||||| |||
Db       1921 NRLIAFASIGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCAGSW 1980

Qy       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040
              | |:||||||||||||||||||||||||||:|||||||||||||||||||||||||:||
Db       1981 LSDMWDWVCTILTDFKNWLTSKLFPKMPGLPFISCQKGYKGVWAGTGIMTTRCPCGASIS 2040

Qy       2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100
              ||||||||||||||||||||||||||||||:|||||||| ||||||||||||||||||
Db       2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCMPKPAPNFKTAIWRVAASEYAEVTQH 2100

Qy       2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160
              | | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 GPYSYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160

Qy       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAEAAARRLARGSPPSEASSSASQLSAPSLRATCT 2220

Qy       2221 THGKAYDVDMVDANLFMGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYMLPKKR 2280
              |||| ||:|||:|||||||||||| |||||||| |||||||||||||||||||| :
Db       2221 THGKTYDMDMVEANLFMGGDVTRIESESKVVVLDSLDSMVEERSDLEPSIPSEYMLPKTK 2280

Qy       2281 FPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLSEDSI 2340
              |||||||||||||||||||||||||||||| ||||||||:|||||||:|||||||| :|
Db       2281 FPPALPAWARPDYNPPLVESWKRPDYQPPTVAGCALPPPKKTPTPPPRKRRTVGLSESTI 2340

Qy       2341 GDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTPPDELALSETGSISSMPPLEGELG 2400
              ||||||||||:|||||||||||||||| |||||:|||||  |||||||||||||||| |
Db       2341 GDALQQLAIKTFGQPPPSGDSGLSTGADAADSGGRTPPDESALSETGSISSMPPLEGEPG 2400

Qy       2401 DPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460
              |||||||||| ||| ||   |||||||||||||||||||||| ||||||||||||||||||
Db       2401 DPDLEPEQVELQPPSQGREVAPGSDSGSWSTCSEEDDSVVCCPMSYSWTGALITPCSPEE 2460

Qy       2461 EKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIKLAASKV 2520
              |||||||||||||||||||||:|||||||||||||||||:|||||||||||||||||||||
Db       2461 EKLPINPLSNSLLRYHNKVYCTTSKSASLRAKKVTFDRMQVLDAYYDSVLKDIKLAASKV 2520

Qy       2521 TARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSETPIPTTI 2580
              :|||||:|||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db       2521 SARLLTLEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSQTPIPTTI 2580

Qy       2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPA 2640
              |||||||||||||||||||||||||||||||||||||||||:|||||||||||||||||||
Db       2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDVTQKLPQAVMGASYGFQYSPA 2640

Qy       2641 QRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTE 2700
              |||||||||||:|::|||||||||||||||||||||||||||:|||||||||||||||||
Db       2641 QRVEFLLKAWADKREPMGFSYDTRCFDSTVTERDIRTEESIYQACSLPEEAHTAIHSLTE 2700

Qy       2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPTMLVCGD 2760
              |||||||| |||| ||||||||||||||||||||||||||||||||||||:|||||||||
Db       2701 RLYVGGPMLNSKGQACGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIVAPTMLVCGD 2760
```

```
Qy       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820

Qy       2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDT 2880
             ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db       2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWVRMVLMTHFFSILMAQDT 2880

Qy       2881 LDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2940
             ||||||||||:||||||||||||||||||||||||:||||||||||||||||||||||||
Db       2881 LDQNLNFEMYGSVYSVSPLDLPAIIERLHGLDAFSLHTYSPHELTRVASALRKLGAPPLR 2940

Qy       2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000
             ||||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||
Db       2941 AWKSRARAVRASLISRGGRAAICGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000

Qy       3001 DIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
              ||||||||||||||  |||||||||| |||||
Db       3001 DIYHSVSRARPRLLLLSLLLLFVGVGLLLLPAR 3033


RESULT 9
Q9IZA6
ID   Q9IZA6      PRELIMINARY;      PRT;  3033 AA.
AC   Q9IZA6;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MD2a-1;
RA   Itakura J., Nagayama K., Enomoto N., Kurosaki M., Watanabe H.,
RA   Sato C.;
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF238481; AAF59940.1; -.
DR   HSSP; Q8JYS1; 1CWX.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR011545; DEAD/DEAH_N.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004109; Peptidase_S29.
DR   InterPro; IPR009003; Pept_Ser_Cys.
DR   InterPro; IPR002518; Pept_U39_HCV_NS2.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00271; Helicase_C; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; Transmembrane.
SQ   SEQUENCE   3033 AA;  329445 MW;  9D0F544E06CA2B28 CRC64;

  Query Match           95.2%;  Score 15347;  DB 2;  Length 3033;
  Best Local Similarity   94.3%;  Pred. No. 0;
  Matches 2860;  Conservative   82;  Mismatches   91;  Indels    0;  Gaps    0;

Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              ||||||||||||||||||:||||||||||||||||||||||||||||||||||||||||
Db          1 MSTNPKPQRKTKRNTNRRPQNVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
              |||||| |||| :||| |  |||||||:|||||||:| |||||||||||||| | |:|||
Db        181 LLSCITIPVSAVQVKNTSNSYMVTNDCSNDSITWQVQDAVLHVPGCVPCEKVENTSRCWI 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
              ||||||||:: ||||| ||||||||||||||||||||||||||||||||||: |||
Db        241 PVSPNVAVRQRGALTQNLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPERHWF 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
              ||:|||||||||||||||||||||||||||| |||||||||:|||||||:||||||||||
Db        301 VQECNCSIYPGTITGHRMAWDMMMNWSPTTTMILAYAMRIPEVIIDIMSGAHWGVMFGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
              |||||||||||||||| ||||| | ||||:  ::    ||||  ||:|:|||:||||||
Db        361 YFSMQGAWAKVVVILLLTAGVDAETQTVGGATGRSVLGFTSLFTRGPQQQIQLINTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
              |||||||||||:|||||||| ::|||||||:||:|||:|||||:||| |||||||||
Db        421 HINRTALNCNDSLNTGFIASLFYYNNFNSSGCPQRMAACRTIEAFRIGWGTLQYEDNVTN 480

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
              |||||||||||:|||:| |::||||||||||||||||:|| ||||||||||||||||||
Db        481 PEDMRPYCWHYPPKQCGIVPARSVCGPVYCFTPSPVVVGTTDKLGVPTYTYWGENETDVFL 540

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
              ||||||| ||||||||||:|:|||||||||||||||||||||||||||||: ||:|
Db        541 LNSTRPPRGSWFGCTWMNSTGFTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPEATYIK 600

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
              ||||||||||:|||||||||||:|||:||:|||||||||||||||||||:|||
Db        601 CGSGPWLTPRCLVDYPYRLWHYPCTINYTVFKVRMYVGGVEHRLTAACNFTRGDRCDLED 660

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
              ||||| :|||||||||||||||:|||||||||||||:|||||||||:|||
Db        661 RDRSQQTPLLHSTTEWAILPCSYSELPALSTGLLHLHQNIVDVQYMYGLSPALTKYVVRW 720

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
              |||:|||||||||||:|||||||||||||||:|||||||||||||||||||||||||
Db        721 EWVVLLFLLLADARVCACVWMLILLGQAEAALEKLVVLHAASAASCNGFLYFVIFFVAAW 780

Qy        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
              ||||| |||| ||||||| | ||:|||||||||||||||| || ||::||||||||||
```

```
Db        781 YIKGRAVPLAAYSLTGLWPFCLLVLALPQQAYAYDASVHGQXGAVLLILITLFTLTPGYK 840

Qy        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
              ||||| |||||||||||||||||| |||| |||||||||  || ||||||||||||||||
Db        841 TLLSRCLWWLCYLLTLGEAMVQEWVPPMQARGGRDGIIWAATIFCPGVVFDITKWLLAVL 900

Qy        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
              || |||:|||||||||||||||||||: |||||||||||||||||||||||||||||||||
Db        901 GPGYLLRGALTRVPYFVRAHALLRMCTVVRHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960

Qy        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
              |||||:|||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db        961 SDWAANGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREILLGPAD 1020

Qy       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080
              ||||||| ||||||||||||||||| ||||||||||||||:|||||||||||||||:|
Db       1021 GYTSKGWRLLAPITAYAQQTRGLLGAIVVSMTGRDKTEQAGEIQILSTVTQSFLGTSIAG 1080

Qy       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140

Qy       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200
              |||||||||||||||||||||||||||||||||||||||||:||||||||||||||||||
Db       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGIFRAAVCSRGVAKSIDFI 1200

Qy       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260

Qy       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGEPITYSTYGKFLADGGCAGGAYDIIICD 1320

Qy       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380
              |||||||:|||||||||||||||||||||||||||||||||||||||||| ||||||||
Db       1321 ECHAVDATTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGSEGEIPFY 1380

Qy       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440
              |:|||||||||||||||||||||||||||||||||:||||||||:|||||||||||||||
Db       1381 GKAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNAVAYYRGLDVSIIPTQGDVVVVA 1440

Qy       1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500
              |||||||||||||:|||||||||||||||||||||| ||||||||||||||||||||||
Db       1441 TDALMTGYTGDFDTVIDCNVAVTQVVDFSLDPTFTITTQTVPQDAVSRSQRRGRTGRGRL 1500

Qy       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db       1501 GIYRYVSTGERASGMFDSVVLGECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560

Qy       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLVAYQATVCARAKAPPPSWDVMWKCLIRLKP 1620

Qy       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680

Qy       1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TGCVSIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740

Qy       1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
              ||||||| |||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 QQASKQAHDIQPAVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800

Qy       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860

Qy       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920

Qy       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCAGSW 1980

Qy       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040
              |||||||||||||||||||||||||||||||:||||||:||||||||||||||||||||
Db       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFISCQKGYRGVWAGTGIMTTRCPCGANIS 2040

Qy       2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100
              |||||||||||||||||| |||||||||||||||||||||||||||| ||||||||||||||
Db       2041 GNVRLGSMRITGPKTCMNTWQGTFPINCYTEGQCVPKPAPNFKTAIWRVAASEYAEVTQH 2100

Qy       2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160
              ||| |||||||||:|||||||||||||||||||||||||||||||||||||| ||||||||
Db       2101 GSYSYITGLTTDNVKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFSVGLNSFVV 2160

Qy       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAEAAARRLARGSPPSEASSSASQLSAPSLRATCT 2220

Qy       2221 THGKAYDVDMVDANLFMGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYMLPKKR 2280
              |||:|:|||||||||||||||||||||||||||||| |||||||||||:||| |
Db       2221 THGRAHDVDMVDANLFMGGDVTRIESESKVVVLDSLDPMAEERSDHEPSIPSEYLLPKSR 2280

Qy       2281 FPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLSEDSI 2340
              |||||||||:|||||||| ||| |||| |||||||||:||||||||: |||||:| :|
Db       2281 FPPALPAWARPEYNPPLVEPWKRADYQPPAVAGCALPPPKKTPTPPPRKPRTVGLNESTI 2340

Qy       2341 GDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTPPDELALSETGSISSMPPLEGELG 2400
              |||||||||:|||||||||||||||| ||||| |||||||| |||||||||||||||| |
Db       2341 ADALQQLAIKTFGQPPPSGDSGLSTGADAADSGGQTPPDELAHSETGSISSMPPLEGEPG 2400

Qy       2401 DPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460
              |||||  ||| : |||||  ||||||||||||||||||||||||||||||||||||||||
Db       2401 DPDLESGQVELESPPQGGEVAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460

Qy       2461 EKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIKLAASKV 2520
              |||||||||||||||||||| |||||||||||||||||::| ||||||||||||||||
Db       2461 EKLPINPLSNSLLRYHNKVYCTTAKSASLRAKKVTFDRMQVLDAHYHSVLKDIKLAASKV 2520

Qy       2521 TARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSETPIPTTI 2580
              :|||||:|||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db       2521 SARLLTLEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSQTPIPTTI 2580

Qy       2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPA 2640
              |||||||||||||||||||||||||||||||||||||||:|||||||||||||||||||
Db       2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDVTQKLPQAVMGASYGFQYSPA 2640

Qy       2641 QRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTE 2700
              |||||||||||||||||||||||||||||||||||||||||||||||| |||||||| ||||||||
Db       2641 QRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYLACSLPEEARTAIHSLTE 2700

Qy       2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPTMLVCGD 2760
              |||||||||||||||||||||||||||||||||||||||||||||||:|||||||||||
Db       2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIVAPTMLVCGD 2760

Qy       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820

Qy       2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDT 2880
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db       2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWVRMVLMTHFFSILMAQDT 2880

Qy       2881 LDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2940
              ||||||||||:||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2881 LDQNLNFEMYGSVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2940

Qy       2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000
              |||||||||||||| |||||:|||||||||||||||||||||||||||||||||||||||
Db       2941 TWKSRARAVRASLISHGGRAAICGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000
```

```
Qy       3001 DIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
              ||||||||||||||||| |||| ||||||||||
Db       3001 DIYHSVSRARPRLLLFSLLLLSVGVGLFLLPAR 3033
```

RESULT 10
Q99IB2
ID Q99IB2 PRELIMINARY; PRT; 3032 AA.
AC Q99IB2;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21316767; PubMed=11424123; DOI=10.1002/jmv.1055;
RA Kato T., Furusaka A., Miyamoto M., Date T., Yasui K., Hiramoto J.,
RA Nagayama K., Tanaka T., Wakita T.;
RT "Sequence analysis of hepatitis C virus isolated from a fulminant
RT hepatitis patient.";
RL J. Med. Virol. 64:334-339(2001).
DR EMBL; AB047645; BAB32878.1; -.
DR HSSP; Q8JYS1; 1CWX.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR GO; GO:0019079; P:viral genome replication; IEA.
DR GO; GO:0019087; P:viral transformation; IEA.
DR InterPro; IPR000345; CytC_heme_BS.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR004109; Peptidase_S29.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR InterPro; IPR002518; Pept_U39_HCV_NS2.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSvir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00998; Viral_RdRP; 1.
DR SMART; SM00487; DEXDc; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
SQ SEQUENCE 3032 AA; 329392 MW; 1FE2A9DBB3442677 CRC64;

Query Match 95.1%; Score 15324.5; DB 2; Length 3032;

Best Local Similarity 94.1%; Pred. No. 0;
Matches 2853; Conservative 81; Mismatches 98; Indels 1; Gaps 1;

```
Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPTDPRHRSRNVG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
              ||||||||||||||||||||||||||||||||||||||||||:||||||||||||||||||
Db        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
              |||||||||||||:||| |  ||||||||:|||||||||| ||||||||||:||| |:|||
Db        181 LLSCITTPVSAAQVKNTSDIYMVTNDCSNDSITWQLQAAALHVPGCVPCERVGNTSRCWI 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
              ||||||||::||||||||||||||||||||||||||||||||||||||||:|||||||||
Db        241 PVSPNVAVRQPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIISPQHHWF 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
              ||:|||||||||||||||||||||||| |||||| |||||||||| ||||||||||||||
Db        301 VQECNCSIYPGTITGHRMAWDMMMNWSPTTTMILAYVMRVPEVIIDIIGGAHWGVMFGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
              |||||||||||||||||||||||||| | | | |     |  :  ||  ||:| |||:||||||
Db        361 YFSMQGAWAKVVVILLLAAGVDAHTRT-GSSVGYATSGIVGLFTSGPKQNIQLINTNGSW 419

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
              |||||||||||||:|||| |||| :|||:|||||:|||| || ||:||| |||||||||
Db        420 HINRTALNCNDSLNTGFIVSLFYARNFNSTGCPERLSACRGIEGFRIGWGTLQYEDNVTN 479

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
              |||||||||||||:|||:| | :|||||||||||||||||||| ||||||||||||||||
Db        480 PEDMRPYCWHYPPKQCGIVPAGSVCGPVYCFTPSPVVVGTTDRLGVPTYTWGENETDVFL 539

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
              |||||||:|||||||||||:|:|||||||||| |||||||||||||||||||||||: ||:|
Db        540 LNSTRPPVGSWFGCTWMNSTGFTKTCGAPPCRIRADFNASTDLLCPTDCFRKHPEATYIK 599

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
              ||||||||||:|||||||||||:|||||||||||||||||||||||||||||||||||||
Db        600 CGSGPWLTPRCLVDYPYRLWHYPCTVNYSIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 659

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
              ||||||:|||||||||||||:|||||||||||||||||||||:||||||||||:|||
Db        660 RDRSQLTPLLHSTTEWAILPCTYSDLPALSTGLLHLHQNIVDVQYMYGLSPALTKYVVRW 719

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
              |||:|||||||||||||:|||||||||||||||:||||| ||||||||||||||||||||
Db        720 EWVVLLFLLLADARVCACVWMLILLGQAEAALEKLVVLHAASTASCNGFLYFVIFFVAAW 779

Qy        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
              ||||| |||||||||||| | |||||||||||||||||:|  ||::|||||||||||||
Db        780 YIKGRAVPLATYSLTGLWPFCLLLLALPQQAYAYDASVHGQVGVVLLILITLFTLTPGYK 839

Qy        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
              ||||| ||||||:|||||||||||||| || |||||||||  || ||:|||||||||||||
Db        840 TLLSRCLWWLCYILTLGEAMVQEWAPSMQARGGRDGIIWAATIFCPGIVFDITKWLLAVL 899

Qy        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
              ||||||: |||||||||||||||||| |||||||||||||||||||||||||||||||||
Db        900 GPAYLLRDALTRVPYFVRAHALLRMCTMVRHLAGGRYVQMALLALGRWTGTYIYDHLTPM 959

Qy        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
              |||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db        960 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREILLGPAD 1019

Qy       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080
              ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||:|
```

```
Db       1020 GYTSKGWRLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSIAG 1079

Qy       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||:||||||||
Db       1080 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGSVDLYLVTRN 1139

Qy       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200
              ||||| ||||||||||||||||||||||||||||||||||||:||||||||||||||||
Db       1140 ADVIPVRRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGIFRAAVCSRGVAKSIDFI 1199

Qy       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1200 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1259

Qy       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320
              ||||||||||||||||||||||||||||||||  ||||||||||||||:||||||||||
Db       1260 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGEAITYSTYGKFLADGGCSGGAYDIIICD 1319

Qy       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1320 ECHATDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1379

Qy       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440
              |:|||||||||||||||||||:||| ||| |||||:||||||||||:||||||||||||
Db       1380 GKAIPLSYIKGGRHLIFCHSKKKCEELAVALRAMGLNAVAYYRGLDVSIIPTQGDVVVVA 1439

Qy       1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500
              ||||||||||||||||||||:|||||||||||| ||||||||||||||||||||||||||
Db       1440 TDALMTGYTGDFDSVIDCNVAVTQVIDFSLDPTFTITTQTVPQDAVSRSQRRGRTGRGRL 1499

Qy       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560
              | ||||||||||||||||||||||||||||||||:||||||||||||||||||||||||
Db       1500 GTYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPAETTVRLRAYFNTPGLPVCQDHLE 1559

Qy       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
              |||  || ||||||||||||||:||||||| ||||||||||||||||||:||||||||||
Db       1560 FWEGAFTDLTHIDAHFLSQTKQAGENFAYLVAYQATVCARAKAPPPSWDIMWKCLTRLKP 1619

Qy       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680
              ||||||||||||||||||||||||||||||||||||||||||||||:||||||||||||
Db       1620 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWILAGGVLAAVAAYCLA 1679

Qy       1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740
              |||| ||||||:|||||||||||| ||||||||||||||||||||||||||||||||||
Db       1680 TGCVSIIGRLHVNQRAVVAPDKEVPYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1739

Qy       1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
              ||||||||:||| || ||||||||||||||||||||||||||||||||||||||||||||
Db       1740 QQASKQAQNIQPAVQTSWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1799

Qy       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860
              ||||||||||||||||||||||||||||||||||||||||||:||||||||||||||||
Db       1800 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAIGSIGLGKVLVDILAGYG 1859

Qy       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1860 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1919

Qy       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db       1920 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCAGSW 1979

Qy       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040
              ||||||||:|||||||||||||||||||:|||||||||||||||||||||||||||| ||
Db       1980 LRDVWDWVCSILTDFKNWLTSKLFPKMPGLPFISCQKGYKGVWAGTGIMTTRCPCGAVIS 2039

Qy       2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100
              |||||||||||||||| |||||||||||||||||||||||| ||||||||||||||||||
Db       2040 GNVRLGSMRITGPKTCMNTWQGTFPINCYTEGQCVPKPAPNFKTAIWRVAASEYAEVTQH 2099

Qy       2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160
              ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2100 GSYSYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2159

Qy       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220
              |||||||||||||||||||||||||||||||| |||||||||||||||||||||| |||||||
Db       2160 GSQLPCDPEPDTDVLMSMLTDPSHITAEAAARRLARGSPPSEASSSASQLSAASLRATCT 2219

Qy       2221 THGKAYDVDMVDANLFMGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYMLPKKR 2280
              |||| ||||||||:||| |||||||| ||||||| ||||||| || ||||||||||: |
Db       2220 THGKTYDVDMVDASLFMRGDVTRIESESKVVVLDYLDPMVEETSDREPSIPSEYMLPRNR 2279

Qy       2281 FPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLSEDSI 2340
              |||||||||||||||||||||||||||||||| ||||||||||:|||||||||||||| :|
Db       2280 FPPALPAWARPDYNPPLVESWKRPDYQPPTVAGCALPPPKKTPTPPPRRRRTVGLSESTI 2339

Qy       2341 GDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTPPDELALSETGSISSMPPLEGELG 2400
              ||||||||:|||||||||||||||||||| ||||| || ||||||||||||||:|||||| |
Db       2340 GDALQQLAVKSFGQPPPSGDSGLSTGADAADSGGQTSPDELALSETGSISSIPPLEGEPG 2399

Qy       2401 DPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460
              |||||||||| ||||||| ||||||||||||||||||:|||||||||||||||||||||
Db       2400 DPDLEPEQVELQPPPQGGEVAPGSDSGSWSTCSEEDDAVVCCSMSYSWTGALITPCSPEE 2459

Qy       2461 EKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIKLAASKV 2520
              |||||||||||||||||||||:||||||||||||||||||:||||||||||||||||||
Db       2460 EKLPINPLSNSLLRYHNKVYCTTSKSASLRAKKVTFDRMQVLDTYYDSVLKDIKLAASKV 2519

Qy       2521 TARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSETPIPTTI 2580
              :|||||:|||||||||||||||||||||||||||: |||||||||:||||||:|||||||
Db       2520 SARLLTLEEACQLTPPHSARSKYGFGAKEVRSLTRRAVNHIKSVWEDLLEDSQTPIPTTI 2579

Qy       2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPA 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db       2580 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDVTQKLPQAVMGASYGFQYSPS 2639

Qy       2641 QRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTE 2700
              |||||||||||||| ||||||||||||||||||:||||||:|||||||| ||||||||
Db       2640 QRVEFLLKAWAEKKGPMGFSYDTRCFDSTVTERDIKTEESIYQACSLPEEARTAIHSLTE 2699

Qy       2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPTMLVCGD 2760
              |||||||||||||:||||||||||||||||||||||||||||||||:|||||||||
Db       2700 RLYVGGPMFNSKGQSCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIVAPTMLVCGD 2759

Qy       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2760 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2819

Qy       2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDT 2880
              |||||||||| ||||||| |||||||||||||||||||||| ||||||||||||||||||
Db       2820 RRRYYLTRDPATPIARAARETVRHSPVNSWLGNIIQYAPTIWVRMVLMTHFFSILMAQDT 2879

Qy       2881 LDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2940
              ||||||||||:|||||||||||||||||| ||||||||||||||||||||||||||||||
Db       2880 LDQNLNFEMYGSVYSVSPLDLPAIIERLHGLGAFSLHTYTPHELTRVASALRKLGAPPLR 2939

Qy       2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000
              |||||||||||||||||||||||||||||||||||:|||||||||||||||||||||||||
Db       2940 TWKSRARAVRASLISRGGRAAVCGRYLFNWAVRTKLKLTPLPEARLLDLSSWFTVGAGGG 2999

Qy       3001 DIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
              |||||||||||||| ||||||||||||||||
Db       3000 DIYHSVSRARPRLLLLSLLLLFVGVGLFLLPAR 3032
```

RESULT 11
Q9IZA3
ID Q9IZA3 PRELIMINARY; PRT; 3033 AA.
AC Q9IZA3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.

```
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MD2a-5;
RA   Itakura J., Nagayama K., Enomoto N., Kurosaki M., Watanabe H.,
RA   Sato C.;
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF238484; AAF59943.1; -.
DR   PIR; S35631; S35631.
DR   HSSP; Q8JYS1; 1CWX.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR011545; DEAD/DEAH_N.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR004109; Peptidase_S29.
DR   InterPro; IPR009003; Pept_Ser_Cys.
DR   InterPro; IPR002518; Pept_U39_HCV_NS2.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; Transmembrane.
SQ   SEQUENCE   3033 AA;  329516 MW;  E016D0E175644593 CRC64;

  Query Match             95.0%;  Score 15322;  DB 2;  Length 3033;
  Best Local Similarity   94.2%;  Pred. No. 0;
  Matches 2858;  Conservative   79;  Mismatches   96;  Indels    0;  Gaps    0;

Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||| || |||||||||
Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRTTRNTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              |||||||||||||||||||||||||||||| ||||||||||||||| ||||||||||
Db         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLCWAGWLLSPRGSRPSWGPTDPRHRSRNVG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
              ||| |||||||||||||| ||||||||||||||||:||||||||||||||||
Db        121 KVIYTLTCGFADLMGYIPVVSAPLGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
              |||||| |||||:|:|||| |||||||:||||||||:||||||||||||||:||| |:|||

Db        181 LLSCITVPVSAAQVRNISTSYMVTNDCSNDSITWQLKAAVLHVPGCVPCERVGNTSRCWI 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
              |||||||| |||||||||||||||||| ||||||||||||||||||:||:||:||||
Db        241 PVSPNVAVQHRGALTQGLRTHIDMVVMSATFCSALYVGDLCGGVMLAAQLFIISPEHHWF 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
              ||:|||||||||||||||||||||||||||||||||||||||||||||||:|||||||||||||
Db        301 VQECNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIINIISGAHWGVMFGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
              |||||||||||| |||||||||| ||:|||||| ||      ||   | :| |||:||||||
Db        361 YFSMQGAWAKVVAILLLAAGVDATTHSVGGSAAHTTRSFVGLFSAGAQQNIQLINTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
              |||||||||||||:|||||||||| |||||||||:||||:|||||:||| |||||:|||
Db        421 HINRTALNCNDSLNTGFIASLFYTHRFNSSGCPERLSACRNIEAFRIGWGTLQYEDDVTN 480

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
              |||||||||||||:|||:| |::|||||||||||||||||| | |||||||||||||||
Db        481 PEDMRPYCWHYPPKQCGIVPARSVCGPVYCFTPSPVVVGTTDRRGVPTYTYWGENETDVFL 540

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
              ||||||| |||||||||:|:|||||||||| |||||||||||||||||: || :
Db        541 LNSTRPPSGSWFGCTWMNSTGFTKTCGAPPCRIRADFNASTDLLCPTDCFRKHPEATYTR 600

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
              |||||||||:||:|||||||||||||||||||||||||||||||||||:|||||||||
Db        601 CGSGPWLTPKCLVDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFSRGDRCNLED 660

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
              |||||||||||||||||||:|||||||||||||||||||:||||||||||:|||
Db        661 RDRSQLSPLLHSTTEWAILPCTYSDLPALSTGLLHLHQNIVDVQYMYGLSPALTKYVVRW 720

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
              |||:|||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db        721 EWVVLLFLLLADARVCACVWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780

Qy        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
              ||||| |||| ||||||| | |||||||||||||||||||||||||:||||||||||
Db        781 YIKGRAVPLAAYSLTGLWPFCLLLLALPQQAYAYDASVHGQIGAALLVLITLFTLTPGYK 840

Qy        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
              ||||| |||||||||||||||||| || |||||||||  || ||:||||||||||
Db        841 TLLSRCLWWLCYLLTLGEAMVQEWAPSMQARGGRDGIIWAATIFCPGIVFDITKWLLAVL 900

Qy        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
              || |||: |||||||||||||||||: ||||||:|||||||||||||||||||||||||
Db        901 GPGYLLRDALTRVPYFVRAHALLRMCTVVRHLAGGKYVQMALLALGRWTGTYIYDHLTPM 960

Qy        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
              |||||:|||||||||||||||||||||||||||||||||||:|||||||||||:||||||
Db        961 SDWAANGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILYGLPVSARLGREILLGPAD 1020

Qy       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080
              ||||||| |||||||||||||||||||||||||||||||||||||||||||:|||
Db       1021 GYTSKGWRLLAPITAYAQQTRGLLGAIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTTISG 1080

Qy       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
              :|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ILWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140

Qy       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200
              |||||||||||||||||||||||||||||||||||||||:|||||||||||||||||
Db       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGIFRAAVCSRGVAKSIDFI 1200

Qy       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260
              ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 PVEALDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260

Qy       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320
              |||||||||| |||||||||||||:|||||  |||||||||||||||||||||||||||||
Db       1261 SVAATLGFGACLSKAHGINPNIRTGVQTVTTGBAITYSTYGKFLADGGCAGGAYDIIICD 1320
```

```
Qy       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380
Db       1321 ECHAVDATTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGHEGEIPFY 1380

Qy       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440
Db       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAGLPGMGLNAVAYYRGLDVSIIPTQGDVVVVA 1440

Qy       1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500
Db       1441 TDALMTGYTDDFDSVIDCNVAVTQVVDFSLDPTFTITTQTVPQDAVSRSQRRGRTGRGRL 1500

Qy       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560
Db       1501 GTYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPGETTVRLRAYFNTPGLPVCQDHLE 1560

Qy       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
Db       1561 FWEAVFTGLTHIDAHFLSQTKQAGENYAYLVAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620

Qy       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680
Db       1621 TLVGPKPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680

Qy       1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740
Db       1681 TGCVSIIGRLHINQRAVVAPDKEVLYEAFDEMEECASKAALIEEGQRIAEMLKSKIQGLL 1740

Qy       1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
Db       1741 QQASKQAQDIQPAVQASWPKVEQFWAKHMWNFHSGIPYLARLSQLPGNPAVASMMAFKCA 1800

Qy       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860
Db       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860

Qy       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920
Db       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920

Qy       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980
Db       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCAGSW 1980

Qy       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040
Db       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFISCQKGYRGVWAGTGIMTTRCPCGANIS 2040

Qy       2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100
Db       2041 GNVRLGSMRITGPKTCMNTWQGTFPINCYTEGQCVPKPAPNFKTAIWRVAASEYAEVTQH 2100

Qy       2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160
Db       2101 GSYSYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160

Qy       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220
Db       2161 GSQLPCDLEPDTDVLMSMLTDPSHITAEAAARRLARGSPPSEASSSASQLSAPSLRATCT 2220

Qy       2221 THGKAYDVDMVDANLFMGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYMLPKKR 2280
Db       2221 THGKTYDVDMVDANLFMGGDVTRIESESKVVVLDSLDPMAEDRSDHEPSIPSEYMLPKTK 2280

Qy       2281 FPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLSEDSI 2340
Db       2281 FPPALPAWARPDYNPPLVESWKRPDYQPPTVAGCALPPPKKTPTPPPRRRRTVGLNKSTI 2340

Qy       2341 GDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTPPDELALSETGSISSMPPLEGELG 2400
Db       2341 GNFLQQLAIKSFGQPPPSGDSGLSTGADAANSGGRTSPDDLALSETGSISSMPPLEGEPG 2400

Qy       2401 DPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460
Db       2401 DPDLEPEQVELQPPPQGGEVAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460

Qy       2461 EKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIKLAASKV 2520
Db       2461 EKLPINPLSNSLLRYHNKVYCTTSKSASLRAKKVTFDRTQVLDAHYDSVLKDIKLAASKV 2520

Qy       2521 TARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSETPIPTTI 2580
Db       2521 NARLLTVEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSQTPIPTTI 2580

Qy       2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPA 2640
Db       2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDVTQKLPQAVMGASYGFQYSPA 2640

Qy       2641 QRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTE 2700
Db       2641 QRVDFLLKAWAEKRDPMGFSYDTRCFDSTVTERDIRTEESIYQACSLPEEARTAIHSLTE 2700

Qy       2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPTMLVCGD 2760
Db       2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIVAPTMLVCGD 2760

Qy       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820
Db       2761 DLVIISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820

Qy       2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDT 2880
Db       2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTVWARMVLMTHFFSILMAQDT 2880

Qy       2881 LDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2940
Db       2881 LDQNLNFEMYGSVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2940

Qy       2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000
Db       2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000

Qy       3001 DIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
Db       3001 DIYHSVSRARPRLLLLSLLLLSVGVGLFLLPAR 3033
```

RESULT 12
Q9IZA2
ID Q9IZA2 PRELIMINARY; PRT; 3033 AA.
AC Q9IZA2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MD2a-7;
RA Itakura J., Nagayama K., Enomoto N., Kurosaki M., Watanabe H.,
RA Sato C.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF238485; AAF59944.1; -.
DR HSSP; Q8JYS1; 1CWX.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.

```
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR011545; DEAD/DEAH_N.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR004109; Peptidase_S29.
DR   InterPro; IPR009003; Pept_Ser_Cys.
DR   InterPro; IPR002518; Pept_U39_HCV_NS2.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; Transmembrane.
SQ   SEQUENCE   3033 AA;  329532 MW;  BE13557BFABA1E87 CRC64;

  Query Match             95.0%;  Score 15321;  DB 2;  Length 3033;
  Best Local Similarity   94.0%;  Pred. No. 0;
  Matches 2852;  Conservative   88;  Mismatches   93;  Indels    0;  Gaps    0;

Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRTTRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||:||||||||||||||
Db        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
              |||||| |||| :||| |  |||||||:||||||||||||||||||:||| |:|||
Db        181 LLSCITIPVSAVQVKNTSNSYMVTNDCSNDSITWQLQAAVLHVPGCVPCERVGNTSRCWI 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
              |||||||||: |||||||||||||||||||||||||||||||||||||||||||:||:||||
Db        241 PVSPNVAVQQRGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIISPEHHWF 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
              ||:|||||| |||||||||||||||||| |||||| ||||||||||||||||||||||||
Db        301 VQECNCSIYLGTITGHRMAWDMMMNWSPTTTMILAYVMRVPEVIIDIISGAHWGVMFGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
              |||||||||||||||: ||||| :  :| : :  |    |  | :| |||:||||||
Db        361 YFSMQGAWAKVVVILLMTAGVDAGSTRIGSAVSHNTDSFAGLLSPGSKQNIQLINTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
              |||||||||||:|||:|||||  :|||||||||:|:|||||||:||| |||||||||
Db        421 HINRTALNCNDSLNTGFLASLFYVRNFNSSGCPERLSSCRSIEAFRIGWGTLQYEDNVTN 480

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
              |||||||||||||:|||:| ||:|||||||||||||||||||| ||||||||||||||
Db        481 PEDMRPYCWHYPPQQCGIVPAKSVCGPVYCFTPSPVVVGTTDRLGVPTYTWGENETDVFL 540

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
              ||||||| |||||||||||:|:|||||||||| |||||||||||||||||||||| ||:|
Db        541 LNSTRPPRGSWFGCTWMNSTGFTKTCGAPPCRIRADFNASTDLLCPTDCFRKHPDATYIK 600

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
              |||||||||||:|||||||||||||:||||||||:|||||||||||||||||:|||
Db        601 CGSGPWLTPRCLVDYPYRLWHYPCTVNYSIFKIRMYVGGIEHRLTAACNFTRGDRCDLED 660

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
              |||||||||||||||||||||:|||||||||||||||||:|||||||||||:|||
Db        661 RDRSQLSPLLHSTTEWAILPCSFSDLPALSTGLLHLHQNIVDVQYMYGLSPALTKYVVRW 720

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
              |||:||||||||||||:|||||||||||||||||:|||||||||||||||||||||
Db        721 EWVVLLFLLLADARVCACVWMLILLGQAEAALEKLVVLHAASAASCNGFLYFVIFFVAAW 780

Qy        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
              |||||||||||||||| | |||||||||||||||| ||:| :|||:|||||||| ||
Db        781 YIKGRVVPLATYSLTGLWPFCLLLLALPQQAYAYDASVQGQLGVSLLVLITLFTLTPAYK 840

Qy        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
              ||||| |||||||||||:||| |||| |||||||||| || ||||||||||||:|
Db        841 TLLSRCLWWLCYLLTLGEAMIQEWVPPMQARGGRDGIIWATTIFCPGVVFDITKWLLALL 900

Qy        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
              || |||:|||||||||||||||||| |||||||||||| |||||||||||| ||
Db        901 GPGYLLRGALTRVPYFVRAHALLRMCTMVRHLAGGRYVQMALLAFGRWTGTYIYDHLAPM 960

Qy        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
              |||||:|||||||||||||||||||||||||||||||||||||||||||||:||||||
Db        961 SDWAANGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREILLGPAD 1020

Qy       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080
              ||||||| |||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       1021 GYTSKGWKLLAPITAYAQQTRGLLGAIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080

Qy       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
              ||||||||||||||||||||||||||||||||||||||||||||:||||||||||||
Db       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCSCGAVDLYLVTRN 1140

Qy       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200
              |||||||||||||||||||||||||||||||||||||||:||||||||||||||||
Db       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGIFRAAVCSRGVAKSIDFI 1200

Qy       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260
              ||| |||:|||||||||||||||||| |||||||||||||||||||||||||||||||
Db       1201 PVEALDIVSRSPTFSDNSTPPAVPQTYQVRYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260

Qy       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320
              |||||||||||||||||||||||||||||| |||||||| |||||:||||||||||
Db       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAAITYSTYGKFFADGGCSGGAYDIIICD 1320

Qy       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380
              ||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 ECHAVDATTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380

Qy       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440
              |||:|||||||||||||||||||||||||||||||||:|||||||||||||||||||||
Db       1381 GRALPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNAVAYYRGLDVSVIPTQGDVVVVA 1440

Qy       1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500
              |||||||||:|||||||||||||||||||||||||||| ||||||||||||||||||||
Db       1441 TDALMTGYTGNFDSVIDCNVAVTQVVDFSLDPTFTITTQTVPQDAVSRSQRRGRTGRGRL 1500

Qy       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560
              |||||||||||||||||||||||||||||||||||:|||||||||||||||||||||||
Db       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPAETTVRLRAYFNTPGLPVCQDHLE 1560

Qy       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
```

```
Db       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLVAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620

Qy       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680

Db       1621 TLVGPTPLLYRLGSVTNEVTFTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680

Qy       1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740

Db       1681 TGCVSIIGRLHINQRAVVAPDKEVLYEAFDEMEECASKAALIEEGQRMAEMLKSKIQGLL 1740

Qy       1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800

Db       1741 QQASKQAQDIQPAVEASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800

Qy       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860

Db       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860

Qy       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920

Db       1861 AGISGALVAFKIMSGEKPSMEDVINLLPGILSPGALVVGVICAGILRRHVGPGEGAVQWM 1920

Qy       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980

Db       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCAGSW 1980

Qy       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040

Db       1981 LRDVWDWVCTILTDFKNWLTSKLFPKLPGLPFVSCQKGYRGVWAGTGIMTTRCPCGANIS 2040

Qy       2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100

Db       2041 GNVRLGSMRITGPKTCMNTWQGTFPINCYTEGQCMPKPAPNFKTAIWRVAASEYAEVTRH 2100

Qy       2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160

Db       2101 GAYSYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFTVGLNSFVV 2160

Qy       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220

Db       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAEAAARRLARGSPPSEASSSASQLSAPSLRATCT 2220

Qy       2221 THGKAYDVDMVDANLFMGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYMLPKKR 2280

Db       2221 THGKTYDVDMVDANLFMGGDVTRIESESKVVVLDSLDPMAEERSDLEPSIPSEYMLPRNT 2280

Qy       2281 FPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLSEDSI 2340

Db       2281 FPPALPAWARPDYNPPLVESWRRPDYQPPTVAGCALPPPKKTPTPPPRRRRTVGLSESTI 2340

Qy       2341 GDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTPPDELALSETGSISSMPPLEGELG 2400

Db       2341 GDALQQLAIKTFGQPPPSGDSGLSTGADAADSGGRTPPDELALSETGSTSSMPPLEGEPG 2400

Qy       2401 DPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460

Db       2401 DPDLEPDQVELRPPSQGGEAAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCCPEE 2460

Qy       2461 EKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIKLAASKV 2520

Db       2461 EKLPINPLSNSLLRYHNKVYSTTSKSASLRAKKVTFDRVQVLDAHYDSVLKDIKLAASKV 2520

Qy       2521 TARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSETPIPTTI 2580

Db       2521 SAKLLTLEEACQLTPPHSARSKYGYGAKEVRSLSGRAVNHIKSVWKDLLEDSQTPIPTTI 2580

Qy       2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPA 2640

Db       2581 MAKNEVFCVDPTKGGKKAARLIVYPDLSVRVCEKMALYDVTQKLPQAVMGASYGFQYSPA 2640

Qy       2641 QRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTE 2700

Db       2641 QRVEFLLKAWADKKDPMGFSYDTRCFDSTVTERDIRTEEAIYQACSLPEEARTAIHSLTE 2700

Qy       2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPTMLVCGD 2760

Db       2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIVAPTMLVCGD 2760

Qy       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820

Db       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGLQG 2820

Qy       2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDT 2880

Db       2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWVRMVLMTHFFSILMAQDT 2880

Qy       2881 LDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2940

Db       2881 LDQDLNFEMYGSVYSVSPLDLPAIIERLHGLDAFLLHTYTPHELTRVASALRKLGAPPLR 2940

Qy       2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000

Db       2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000

Qy       3001 DIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033

Db       3001 DIYHSVSRARPRLLLLGLLLLSVGVGLFLLPAR 3033
```

RESULT 13
Q9Q9B0
ID Q9Q9B0 PRELIMINARY; PRT; 3033 AA.
AC Q9Q9B0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21361470; PubMed=11468731; DOI=10.1002/jmv.1073;
RA Kurihara C., Ishiyama N., Nishiyama Y., Fukushi S., Kageyama T.,
RA Katayama K., Miura S.;
RT "Molecular characterization of hepatitis C virus genotype 2a from the
RT entire sequences of four isolates.";
RL J. Med. Virol. 64:466-475(2001).
DR EMBL; AF169002; AAF25610.1; -.
DR HSSP; Q8JYS1; 1CWX.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR GO; GO:0019079; P:viral genome replication; IEA.
DR GO; GO:0019087; P:viral transformation; IEA.
DR InterPro; IPR000345; CytC_heme_BS.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.

```
DR   InterPro; IPR004109; Peptidase_S29.
DR   InterPro; IPR009003; Pept_Ser_Cys.
DR   InterPro; IPR002518; Pept_U39_HCV_NS2.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; Transmembrane.
SQ   SEQUENCE   3033 AA;  329180 MW;  6E3F51FD3A8A4C24 CRC64;

  Query Match           95.0%;  Score 15311;  DB 2;  Length 3033;
  Best Local Similarity   94.1%;  Pred. No. 0;
  Matches 2855;  Conservative   75;  Mismatches  103;  Indels    0;  Gaps    0;

Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRR 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              |||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RRQPIPKDRRSTGRSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
              |||||||||||||||||||||||||||||| ||||||| ||||:||||||||||||||||
Db        121 KVIDTLTCGFADLMGYIPVVGAPLGGVATALAHGVRALEDGINFATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
              |||||| |||||:||| |:||||||| |||||||||| |||||||||| |    |:|||
Db        181 LLSCITVPVSAAQVKNTSSGYMVTNDCANDSITWQLQAVVLHVPGCVPCEXVXXTSRCWI 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
              ||||||||::|||||||||||:||||||||||||||||||||||||||||:||||
Db        241 PVSPNVAVRQPGALTQGLRTHLDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPEHHWF 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
              ||:|||||||| ||| ||||||||||||||||||||||||||||||||||||||||||||
Db        301 VQECNCSIYPGAITGQRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
              |||||||||||||||||||| |   ||:||:    :|||| |  | |||:||||||
Db        361 YFSMQGAWAKVVVILLLAAGVDANTXXXGGAAARNAYGVTSLFSXGAXQNIQLINTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
              |||||||||||||:|||:| ||| : |||||||||:| ||:|||||:||| |||||||||
Db        421 HINRTALNCNDSLNTGFLAXLFYXNRFNSSGCPERLSVCRNIEAFRIGWGTLQYEDNVTN 480

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
              |||||||||||||:|| :| |::||||||||||||||||| | ||||||||||||||||
Db        481 PEDMRPYCWHYPPKQCSIVPARSVCGPVYCFTPSPVVVGTTDRRGVPTYTWGENETDVFL 540

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
              ||||||| ||||||||||:|:||||||||||||||||||||||||||||: ||:|
Db        541 LNSTRPPQGSWFGCTWMNSTGFTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPEATYIK 600

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
              ||||||||||||:||||||||||||||||||||||||||||||||||||||:|||
Db        601 CGSGPWLTPRCLVDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCDLED 660

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
              ||||||||||||||||||||||||||||||||||||||||||:||||||||||||||||
Db        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQYMYGLSPALTKYIVRW 720

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
              |||:|||||||||||||||| ||:||||||||||:|||||||||||||||||||||||||
Db        721 EWVVLLFLLLADARVCACLWMFILMGQAEAALEKLVVLHAASAASCNGFLYFVIFFVAAW 780

Qy        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
              ||||| |||| ||||||||| |||| ||||||||:|||||||||||||:||||||||||
Db        781 YIKGRAVPLAAYSLTGLWSFCLLLLVLPQQAYAYEASVHGQIGAALLVLITLFTLTPGYK 840

Qy        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
              ||||| |||||||||||||||||||| || |||||||||| || |||||||||||||| |
Db        841 TLLSRCLWWLCYLLTLGEAMVQEWAPSMQARGGRDGIIWAATIFCPGVVFDITKWLLAAL 900

Qy        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
              |  |||:|||||||||||||||||||:|:|||||:|||||||||||||||||||||||||
Db        901 GSGYLLRGALTRVPYFVRAHALLRMCTIAKHLAGGKYVQMALLALGRWTGTYIYDHLTPM 960

Qy        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREILLGPAD 1020

Qy       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080
              ||||||| |||||||||||||||| ||||||||||||||||||||||||||||||||||
Db       1021 GYTSKGWRLLAPITAYAQQTRGLLGAIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080

Qy       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140

Qy       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200
              |||||||||||||||||||||||||||||||||||||||||:||||||||||||||||
Db       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGIFRAAVCSRGVAKSIDFI 1200

Qy       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||:||||||||||
Db       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYASQGYKVLVLNP 1260

Qy       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320
              |||||||||||||||||||||||||||||||  |||||||||||||||||||||||||||
Db       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGEAITYSTYGKFLADGGCAGGAYDIIICD 1320

Qy       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380
              ||||||:||||||||||||||||||||||||||||||||||||||||| |||||||
Db       1321 ECHAVDATTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGHEGEIPFY 1380

Qy       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440
              |:|||||||||||||||||||||||:|||||||:|||||||| || ||||||||
Db       1381 GKAIPLSYIKGGRHLIFCHSKKKCDELSAALRGMGLNAVAYYRGLDVSAIPAQGDVVVVA 1440

Qy       1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQTVPQDAVSRSQRRGRTGRGRL 1500

Qy       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560
              |||||||||||||||||||||||||||||||||:||||||||||||||||||||||||
Db       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPAETTVRLRAYFNTPGLPVCQDHLE 1560

Qy       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLVAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620

Qy       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680

Qy       1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TGCVSIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740

Qy       1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 QQASKQAQDIQPAVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800

Qy       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860
```

```
Db       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSVGLGKVLVDILAGYG 1860

Qy       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920
Db       1861 AGISGALVAFKVMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920

Qy       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980
Db       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCAGSW 1980

Qy       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040
Db       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFISCQKGYRGVWAGTGIMTTRCSCGAEIS 2040

Qy       2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100
Db       2041 GNVRLGSMRITGPRTCMNTWQGTFPINCYTEGQCVPKPAPNFKTAIWRVAASEYAEVTQH 2100

Qy       2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160
Db       2101 GPYSYITGLTTDNLKVPCQIPSPEFFSWVDGVQIHRFAPKPKPFFRDEVSFCVGLNSFVV 2160

Qy       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220
Db       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAEAAARRLARGSPPSEASSSASQLSAPSLRATCA 2220

Qy       2221 THGKAYDVDMVDANLFMGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYMLPKKR 2280
Db       2221 THGKTYDVDMVDANLFMGGDVTRIESESKVVVLDSLDPMAEERSDLEPSIPSEYMLPRNR 2280

Qy       2281 FPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLSEDSI 2340
Db       2281 FPPALPAWARPDYNPPLVESWKRPDYQPPTVAGCALPPPKKAPTPPPRRRRTVGLSESTI 2340

Qy       2341 GDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTPPDELALSETGSISSMPPLEGELG 2400
Db       2341 GDALQQLAIKTFGQPPPSGDSGLSTGADAADSGGQASPDELALSETGSTSSMPPLEGEPG 2400

Qy       2401 DPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460
Db       2401 DPDLEPERVELQALPQGGEVAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPSGPEE 2460

Qy       2461 EKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIKLAASKV 2520
Db       2461 EKLPINPLSNSLLRYHNKVYCTTSRTASLRAKKVTFDRTQVLDAYYDSTLKDIKLAASRV 2520

Qy       2521 TARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSETPIPTTI 2580
Db       2521 SARLLTLEEACQLTPPHSARSRYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSQTPIPTTI 2580

Qy       2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPA 2640
Db       2581 MAKNEVFCVEPTKGGKKAARLIVYPDLSVRVCEKMALYDVTQKLPQAVMGASYGFQYSPA 2640

Qy       2641 QRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTE 2700
Db       2641 QRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYQACSLPEEARTAIHSLTE 2700

Qy       2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPTMLVCGD 2760
Db       2701 RLYVGGPMLNSKGQACGYRRCRASGVLTTSMGNTITCYIKALAACKAAGIVAPTMLVCGD 2760

Qy       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820
Db       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820

Qy       2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDT 2880
Db       2821 RRRYYLTRDPTTPXARAAWETVRHSPVNSWLGNIIQYAPTIWVRMVLMTHFYSILMAQDT 2880

Qy       2881 LDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2940
Db       2881 LDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVAAALRKLGAPPLR 2940

Qy       2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000
Db       2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARHLDLSSWFTVGAGGG 3000

Qy       3001 DIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
Db       3001 DIYHSVLRARPRLLLLGLLLLFVGVGLFLLPAR 3033
```

RESULT 14
Q9Q9A8
ID Q9Q9A8 PRELIMINARY; PRT; 3033 AA.
AC Q9Q9A8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21361470; PubMed=11468731; DOI=10.1002/jmv.1073;
RA Kurihara C., Ishiyama N., Nishiyama Y., Fukushi S., Kageyama T.,
RA Katayama K., Miura S.;
RT "Molecular characterization of hepatitis C virus genotype 2a from the
RT entire sequences of four isolates.";
RL J. Med. Virol. 64:466-475(2001).
DR EMBL; AF169004; AAF25612.1; -.
DR HSSP; Q8JYS1; 1CWX.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR GO; GO:0019079; P:viral genome replication; IEA.
DR GO; GO:0019087; P:viral transformation; IEA.
DR InterPro; IPR000345; CytC_heme_BS.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR004109; Peptidase_S29.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR InterPro; IPR002518; Pept_U39_HCV_NS2.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSvir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00998; Viral_RdRP; 1.
DR SMART; SM00487; DEXDc; 1.

```
DR   PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; Transmembrane.
SQ   SEQUENCE   3033 AA;  329175 MW;  74932A91979E78A5 CRC64;

  Query Match           94.7%;  Score 15266;  DB 2;  Length 3033;
  Best Local Similarity  94.0%;  Pred. No. 0;
  Matches 2850;  Conservative   71;  Mismatches  112;  Indels    0;  Gaps    0;

Qy        1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
            ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 MSTKPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Qy       61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
            |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db       61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPTDPRHRSRNVG 120

Qy      121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
            ||||||||||||||||||||||||||||||||||||||||||||:||||||||||||||||
Db      121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180

Qy      181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
            |||||| |||||:||| ||||||||||:||||||||||||||||||||||||| |:|||
Db      181 LLSCITIPVSAAQVKNTSTGYMVTNDCSNDSITWQLQAAVLHVPGCVPCEKVGNXSRCWI 240

Qy      241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300

Qy      301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
            ||:|||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      301 VQECNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIIGGAHWGVMFGLA 360

Qy      361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
            |||||||||||||||||||| | |  |  | |  :  ||  | :|  ||:|:||||
Db      361 YFSMQGAWAKVVVILLLAAGVDAXTXTXAGXXAXTVSGIXXLFXPGAKQNXQLINSNGSW 420

Qy      421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
            |||||||||||||:|||:||||||||||||||:||||:|||||:||| ||||:||||
Db      421 HINRTALNCNDSLNTGFLASLFYTHSFNSSGCPERLSACRNIEAFRIGWGTLQYEENVTN 480

Qy      481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
            |||||||||||||:|||:| |::|||||||||||||||||| ||||||||||||||
Db      481 PEDMRPYCWHYPPKQCGIVPAESVCGPVYCFTPSPVVVGTTDRLGVPTYTWGENETDVFL 540

Qy      541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
            |||||||||||||||:|:|| ||||||| |||||||||||||||||||: ||:|
Db      541 LNSTRPPSGSWFGCTWMNSTGFTKACGAPPCRXRADFNASTDLLCPTDCFRKHPEATYMK 600

Qy      601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
            |||||||||||||:|||||||||||||||||||||||||||||||||:|| ||||||
Db      601 CGSGPWLTPRCLVDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFSRGYRCNLED 660

Qy      661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
            ||| |:|||||||||||||:|||||||:|||||||:||||||||:|||
Db      661 RDRGQVSPLLHSTTEWAILPCSYTDLPALSTGLIHLHQNIVDVQYMYGLSPALTKYVVRW 720

Qy      721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
            |||:|||||||||||||||||||||||||||:|||||||||||||||||||||||||:|||
Db      721 EWVVLLFLLLADARVCACLWMLILLGQAEAALEKLVVLHAASAASCNGFLYFVIFFIAAW 780

Qy      781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
            ||||| |||||||||||| | |||||||||||||| ||||||||::|||||||| ||
Db      781 YIKGRXVPLATYSLTGLWPFCLLLLALPQQAYAYDGSVHGQIGAALLILITLFTLTPAYK 840

Qy      841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
            ||||| |||||||||||||||||| || |||||||||  || ||||||||||||||||
Db      841 TLLSRCLWWLCYLLTLGEAMVQEWAPSMQARGGRDGIIWAATIFCPGVVFDITKWLLAVL 900

Qy      901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
            || |||: ||||||||||||||||||  |||||||||||||||||||||||||||||||||
Db      901 GPGYLLRDALTRVPYFVRAHALLRMCTXVRHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960

Qy      961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
            |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||:||||||
Db      961 SDWAASGLRDLAVAVEPIIFSPMEKKVITWGAETAACGDILHGLPVSARLGREILLGPAD 1020

Qy     1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080
            ||||||| |||||||||||||||| ||||||||||||||||||||||||||||||| |
Db     1021 GYTSKGWRLLAPITAYAQQTRGLLGAIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSIXG 1080

Qy     1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140

Qy     1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200
            |||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||||
Db     1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGIFRAAVCSRGVAKSIDFI 1200

Qy     1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260
            ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1201 PVEALAIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260

Qy     1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320
            ||||||||||||:|||||||||||||||||| |||||||||||||| ||||||||||
Db     1261 SVAATLGFGAYLSRAHGINPNIRTGVRTVTTGAAITYSTYGKFLADGGCXGGAYDIIICD 1320

Qy     1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380
            |||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1321 ECHSVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380

Qy     1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440
            |:|||||||||||||||||||||||||||||||||||:|||||||||||||||||||||||
Db     1381 GKAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNAVAYYRGLDVSVIPTQGDVVVVA 1440

Qy     1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500
            |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db     1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQTVPQDAVSRSQRRGRTGRGRL 1500

Qy     1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560
            |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db     1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPVETTVRLRAYFNTPGLPVCQDHLE 1560

Qy     1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
            |||||||||||||||||||||||||||||||| ||||||||||||||||:||||||||||
Db     1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLVAYQATVCARAKAPPPSWDIMWKCLTRLKP 1620

Qy     1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680

Qy     1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740
            |||| |||||| ||: |||||||||||||||||||||||||||:|||||||||||||||||
Db     1681 TGCVSIIGRLHXNQQVVVAPDKEVLYEAFDEMEECASRAALLEEGQRIAEMLKSKIQGLL 1740

Qy     1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
            |||||||||||| ||||||:||||||||||||||||||||||||||||||||||||||||
Db     1741 QQASKQAQDIQPAVQASWPRVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800

Qy     1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860

Qy     1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920

Qy     1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db     1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCAGSW 1980

Qy     1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040
            |||||||||:|| |||||||||||||||| :|||||||||||||||||||||||||||||
Db     1981 LRDVWDWVCSILIDFKNWLTSKLFPKMPGLPLISCQKGYKGVWAGTGIMTTRCPCGANIS 2040

Qy     2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100
```

```
                ||||||||||||||||||  | ||||||||||||||||||||| ||||||||||||||||
Db       2041 GNVRLGSMRITGPKTCMNTWHGTFPINCYTEGQCVPKPAPNFKTAIWRVAASEYAEVTQH 2100

Qy       2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160
              ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 GSYSYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160

Qy       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220
              |||||||||||||||| |||||||||||| |||||||||||||||||||||||||||||||
Db       2161 GSQLPCDPEPDTDVLTSMLTDPSHITAEAAARRLARGSPPSEASSSASQLSAPSLRATCT 2220

Qy       2221 THGKAYDVDMVDANLFMGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYMLPKKR 2280
              |||| |||||||||||||||||| |||:|||||| |||| || ||||||||||: :
Db       2221 THGKTYDVDMVDANLFMGGDVTRIESESKVIVLDSLDSMVEEGSDHEPSIPSEYMLPRTK 2280

Qy       2281 FPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLSEDSI 2340
              |||||| |||||||||:|||||||||| |||||||| ::|||||||||||||||| :|
Db       2281 FPPALPVWARPDYNPPVVESWKRPDYQPPTVAGCALPPSKETPTPPPRRRRTVGLSESTI 2340

Qy       2341 GDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTPPDELALSETGSISSMPPLEGELG 2400
              ||||||||||:||||| :|||| |||| ||||| :|||||  ||||| |||||||||| |
Db       2341 GDALQQLAIKTFGQPPSTGDSGXSTGADAADSGGRTPPDELTPSETGSTSSMPPLEGEPG 2400

Qy       2401 DPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460
              |||||||||| ||||| |  |||||||||||||||||||||||||||||||||||||||||
Db       2401 DPDLEPEQVELQPPPQEGXVAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460

Qy       2461 EKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIKLAASKV 2520
              ||||||| |||||||||||||:||||||||||||||::|||||||||||||||
Db       2461 EKLPINPXSNSLLRYHNKVYCTTSKSASLRAKKVTFDRMQVLDAHYDSVLKDIKLAASKV 2520

Qy       2521 TARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSETPIPTTI 2580
              :|||||:|||||||||||:|||||||||||||| |||||||||||||||| |||||||
Db       2521 SARLLTVEEACQLTPPHSAKSKYGFGAKEVRSLSRRAVNHIKSVWKDLLEDSGTPIPTTI 2580

Qy       2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPA 2640
              |||||||||:|||||||||||||||||||||||||||||:|||||||||||||||||||||
Db       2581 MAKNEVFCVEPTKGGKKAARLIVYPDLGVRVCEKMALYDVTQKLPQAVMGASYGFQYSPA 2640

Qy       2641 QRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTE 2700
              ||||||||||:|||||||||||||||||||||||||||||:|||||||| ||||||||
Db       2641 QRVEFLLKAWADKKDPMGFSYDTRCFDSTVTERDIRTEESIYQACSLPEEARTAIHSLTE 2700

Qy       2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPTMLVCGD 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||:|||||||||
Db       2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIVAPTMLVCGD 2760

Qy       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820
              |||||||||| |||||||| |||||||||| |||||||||||||||||||  |||
Db       2761 DLVVISESQGAEEDERNLRVFTEAMTRYSAPSGDPPRPEYDLELITSCSSNVSVARDPQG 2820

Qy       2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDT 2880
              |||||||||||||||||||||| |||||||||||||||| ||||||||||||||||||||
Db       2821 RRRYYLTRDPTTPIARAAWETVRHFPVNSWLGNIIQYAPTIWVRMVLMTHFFSILMAQDT 2880

Qy       2881 LDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2940
              ||||||||||:|||||| |||||||||||||||||||||||||||||||||||||||||||
Db       2881 LDQNLNFEMYGSVYSVSPSDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2940

Qy       2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000
              |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db       2941 AWKSRARAVRASLISHGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000

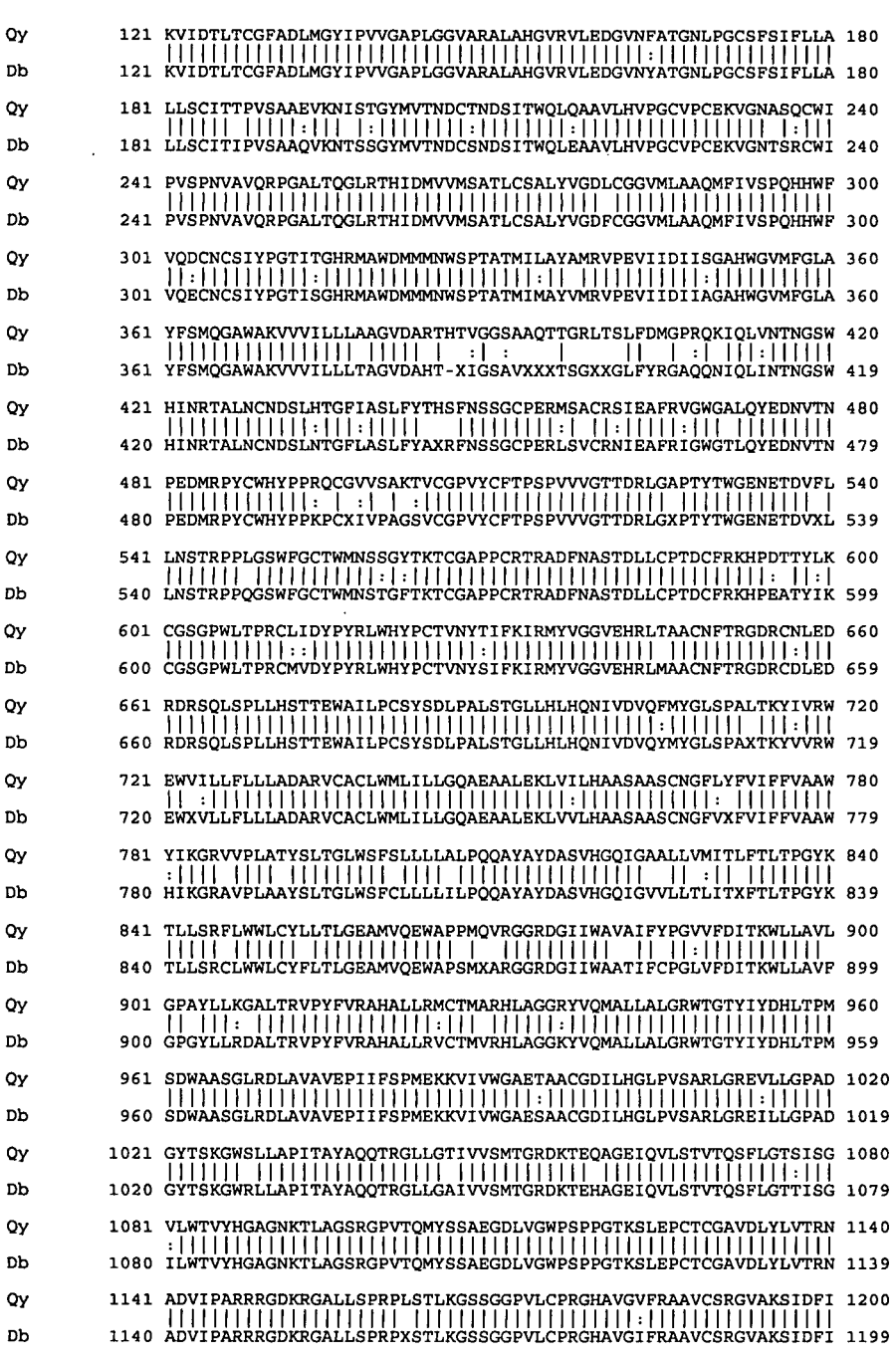
Qy       3001 DIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
              |||  | ||||||||  |||| ||:||||||||
Db       3001 DIYLCVLRARPRLLLLSLLLLSVGLGLFLLPAR 3033


RESULT 15
Q9Q9A9
ID   Q9Q9A9      PRELIMINARY;      PRT;  3032 AA.
AC   Q9Q9A9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Polyprotein.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21361470; PubMed=11468731; DOI=10.1002/jmv.1073;
RA   Kurihara C., Ishiyama N., Nishiyama Y., Fukushi S., Kageyama T.,
RA   Katayama K., Miura S.;
RT   "Molecular characterization of hepatitis C virus genotype 2a from the
RT   entire sequences of four isolates.";
RL   J. Med. Virol. 64:466-475(2001).
DR   EMBL; AF169003; AAF25611.1; -.
DR   HSSP; Q8JYS1; 1CWX.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0019028; C:viral capsid; IEA.
DR   GO; GO:0019031; C:viral envelope; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR   GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006350; P:transcription; IEA.
DR   GO; GO:0019079; P:viral genome replication; IEA.
DR   GO; GO:0019087; P:viral transformation; IEA.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR011545; DEAD/DEAH_N.
DR   InterPro; IPR002522; HCV_capsid.
DR   InterPro; IPR002521; HCV_core.
DR   InterPro; IPR002519; HCV_env.
DR   InterPro; IPR002531; HCV_NS1.
DR   InterPro; IPR000745; HCV_NS4a.
DR   InterPro; IPR001490; HCV_NS4b.
DR   InterPro; IPR002868; HCV_NS5a.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR004109; Peptidase_S29.
DR   InterPro; IPR009003; Pept_Ser_Cys.
DR   InterPro; IPR002518; Pept_U39_HCV_NS2.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF01543; HCV_capsid; 1.
DR   Pfam; PF01542; HCV_core; 1.
DR   Pfam; PF01539; HCV_env; 1.
DR   Pfam; PF01560; HCV_NS1; 1.
DR   Pfam; PF01538; HCV_NS2; 1.
DR   Pfam; PF02907; HCV_NS3; 1.
DR   Pfam; PF01006; HCV_NS4a; 1.
DR   Pfam; PF01001; HCV_NS4b; 1.
DR   Pfam; PF01506; HCV_NS5a; 1.
DR   Pfam; PF00998; Viral_RdRP; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
KW   Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW   Polyprotein; Transmembrane.
SQ   SEQUENCE   3032 AA;  329397 MW;  BAF67F4730C124B9 CRC64;

  Query Match             94.6%;  Score 15249.5;  DB 2;  Length 3032;
  Best Local Similarity   93.9%;  Pred. No. 0;
  Matches 2847;  Conservative   69;  Mismatches  116;  Indels    1;  Gaps    1;

Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              |||||||| ||||||||| |||||||||||||||||||||||||:||| ||||||||||
Db         61 RRQPIPKHRRSTGKSWGTPGYPWPLYGNEGLGWAGWLLSPRGSRPTWGPTDPRHRSRNVG 120
```


Qy 121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
Db 121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180
Qy 181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
Db 181 LLSCITIPVSAAQVKNTSSGYMVTNDCSNDSITWQLEAAVLHVPGCVPCEKVGNTSRCWI 240
Qy 241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
Db 241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDFCGGVMLAAQMFIVSPQHHWF 300
Qy 301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
Db 301 VQECNCSIYPGTISGHRMAWDMMMNWSPTATMIMAYVMRVPEVIIDIIAGAHWGVMFGLA 360
Qy 361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
Db 361 YFSMQGAWAKVVVILLLTAGVDAHT-XIGSAVXXXTSGXXGLFYRGAQQNIQLINTNGSW 419
Qy 421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
Db 420 HINRTALNCNDSLNTGFLASLFYAXRFNSSGCPERLSVCRNIEAFRIGWGTLQYEDNVTN 479
Qy 481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
Db 480 PEDMRPYCWHYPPKPCXIVPAGSVCGPVYCFTPSPVVVGTTDRLGXPTYTWGENETDVXL 539
Qy 541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
Db 540 LNSTRPPQGSWFGCTWMNSTGFTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPEATYIK 599
Qy 601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
Db 600 CGSGPWLTPRCMVDYPYRLWHYPCTVNYSIFKIRMYVGGVEHRLMAACNFTRGDRCDLED 659
Qy 661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
Db 660 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQYMYGLSPAXTKYVVRW 719
Qy 721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
Db 720 EWXVLLFLLLADARVCACLWMLILLGQAEAALEKLVVLHAASAASCNGFVXFVIFFVAAW 779
Qy 781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
Db 780 HIKGRAVPLAAYSLTGLWSFCLLLLILPQQAYAYDASVHGQIGVVLLTLITXFTLTPGYK 839
Qy 841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
Db 840 TLLSRCLWWLCYFLTLGEAMVQEWAPSMXARGGRDGIIWAATIFCPGLVFDITKWLLAVF 899
Qy 901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
Db 900 GPGYLLRDALTRVPYFVRAHALLRVCTMVRHLAGGKYVQMALLALGRWTGTYIYDHLTPM 959
Qy 961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
Db 960 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAESAACGDILHGLPVSARLGREILLGPAD 1019
Qy 1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080
Db 1020 GYTSKGWRLLAPITAYAQQTRGLLGAIVVSMTGRDKTEHAGEIQVLSTVTQSFLGTTISG 1079
Qy 1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
Db 1080 ILWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1139
Qy 1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200
Db 1140 ADVIPARRRGDKRGALLSPRPXSTLKGSSGGPVLCPRGHAVGIFRAAVCSRGVAKSIDFI 1199
Qy 1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260
Db 1200 PVEALDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYASQGYKVLVLNP 1259
Qy 1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320
Db 1260 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGEPITYSTYGKFLADGGCTGGAYDIIICD 1319
Qy 1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380
Db 1320 ECHSVDATTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1379
Qy 1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440
Db 1380 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNAVAYYRGLDVSTIPTQGDVVVVA 1439
Qy 1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500
Db 1440 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQTVPQDAVSRSQRRGRTGRGRL 1499
Qy 1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560
Db 1500 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPVETTVRLRAYFNTPGLPVCQDHLE 1559
Qy 1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
Db 1560 FWEAVFTGLTHIDAHFLSQTKQSGXNFAYLVAYQATVCARAKAPPPSWDVMWKCLTRLKP 1619
Qy 1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680
Db 1620 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1679
Qy 1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740
Db 1680 TGCVSIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLRSKIQGLL 1739
Qy 1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
Db 1740 QQASKQAQDIQPAVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1799
Qy 1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860
Db 1800 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1859
Qy 1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920
Db 1860 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1919
Qy 1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980
Db 1920 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCAGSW 1979
Qy 1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040
Db 1980 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFISCQKGYKGVWAGTGIMTTRCPCGANIS 2039
Qy 2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100
Db 2040 GNVRLGSMRITGPKTCMNTWQGTFPINCYTEGQCVPKPTLNFKTAIWRVAASEYAEVTQH 2099
Qy 2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160
Db 2100 GSYSYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPXPKPFFRDEVSFCVGLNSFVV 2159
Qy 2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220
Db 2160 GSQLPCDPEPDTDVLMSMLTDPTHITAEAAARRLARGSPPSEASSSASQLSAPSLRATCT 2219
Qy 2221 THGKAYDVDMVDANLFMGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYMLPKKR 2280
Db 2220 THGKAYDVDMVDANLFMGGDVTRIESESKVVVLDSLDPMAEEKSDLEPSIPSEYMLPKSK 2279
Qy 2281 FPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLSEDSI 2340

```
Db       2280 FPPALPAWARPDYNPPLVESWKKPDYQPPTVAGCALPPPKKTPTPPPRSRRTVGLRESTI 2339

Qy       2341 GDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTPPDELALSETGSISSMPPLEGELG 2400

Db       2340 GDALQQLAIKTFGQPPPSGDSGLSTGADVADSGGQTPPDELALSETGSISSMPPLEGEPG 2399

Qy       2401 DPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460

Db       2400 DPDLEPELVELQPPPQGGEVAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2459

Qy       2461 EKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIKLAASKV 2520

Db       2460 EKLPINPLSNSLLRYHNKVYCTTSKSASLRAKKVTFDRMQVLDAYYDSVLKDIKLAASKV 2519

Qy       2521 TARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSETPIPTTI 2580

Db       2520 SARLLTLEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSQTPIPTTI 2579

Qy       2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPA 2640

Db       2580 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMPLYDVTQKLPQAVMGASYGFQYSPA 2639

Qy       2641 QRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTE 2700

Db       2640 QRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLXEEAHTAIHSLTE 2699

Qy       2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPTMLVCGD 2760

Db       2700 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIVAPTMLVCGD 2759

Qy       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820

Db       2760 DLVVISESQGTEEDERNLRXFTEAMTRYSAPSGDPPRPEYDLELITSCSSNVSVARDPQG 2819

Qy       2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDT 2880

Db       2820 RRRYYLTRDPTTPIARAAWETVRHFPVNSWLGNIIQYAPTIWVRMVLMTHFFSILMAQDT 2879

Qy       2881 LDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2940

Db       2880 LDQNLNFEMYGSVYSVSPSDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2939

Qy       2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000

Db       2940 AWKSRARAVRASLISHGGRAAVCGRYLFNWAVXTKLKLTPLPEARLLDLSSWFTVGAGGG 2999

Qy       3001 DIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033

Db       3000 DIYLCVLRARPRLLLLSLLLLSVGLGLFLLPAR 3032


Search completed: October 28, 2005, 16:04:36
Job time : 345 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2005, 15:50:12 ; Search time 65 Seconds
(without alignments)
3483.240 Million cell updates/sec

Title: US-09-980-559-2
Perfect score: 16120
Sequence: 1 MSTNPKPQRKTKRNTNRRPQ.....LLFGLLLLFVGVGLFLLPAR 3033

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	15812	98.1	3033	1	US-07-925-695-5	Sequence 5, Appli
2	13966	86.6	3033	1	US-07-925-695-8	Sequence 8, Appli
3	13881	86.1	3033	1	US-07-925-695-9	Sequence 9, Appli
4	12027.5	74.6	3010	4	US-09-539-601-3	Sequence 3, Appli
5	12010.5	74.5	3010	4	US-09-539-601-21	Sequence 21, Appl
6	12001.5	74.5	3010	4	US-09-539-601-27	Sequence 27, Appl
7	11984.5	74.3	3010	1	US-08-324-977-2	Sequence 2, Appli
8	11984.5	74.3	3010	1	US-08-324-977-14	Sequence 14, Appl
9	11984.5	74.3	3010	2	US-08-384-616-2	Sequence 2, Appli
10	11984.5	74.3	3010	2	US-08-384-616-14	Sequence 14, Appl
11	11984.5	74.3	3010	2	US-08-904-686A-2	Sequence 2, Appli
12	11984.5	74.3	3010	2	US-08-904-686A-14	Sequence 14, Appl
13	11984.5	74.3	3010	3	US-09-315-850-2	Sequence 2, Appli
14	11984.5	74.3	3010	3	US-09-315-850-14	Sequence 14, Appl
15	11984.5	74.3	3010	4	US-09-539-601-33	Sequence 33, Appl
16	11973	74.3	3012	3	US-08-811-566-2	Sequence 2, Appli
17	11973	74.3	3012	3	US-09-034-756-2	Sequence 2, Appli
18	11966	74.2	3011	3	US-09-014-416-1	Sequence 1, Appli
19	11966	74.2	3011	4	US-09-952-572-9	Sequence 9, Appli
20	11958.5	74.2	3010	3	US-09-014-416-3	Sequence 3, Appli
21	11947	74.1	3011	1	US-08-453-552-2	Sequence 2, Appli
22	11947	74.1	3011	2	US-08-710-637-2	Sequence 2, Appli
23	11947	74.1	3011	5	PCT-US93-00907-2	Sequence 2, Appli
24	11945	74.1	3011	3	US-09-014-416-5	Sequence 5, Appli
25	11933	74.0	3011	3	US-08-850-328-1	Sequence 1, Appli
26	11922	74.0	3011	1	US-08-440-103-36	Sequence 36, Appl
27	11922	74.0	3011	1	US-08-440-542-36	Sequence 36, Appl
28	11922	74.0	3011	1	US-07-910-760-10	Sequence 10, Appl
29	11922	74.0	3011	1	US-08-440-519-10	Sequence 10, Appl
30	11922	74.0	3011	1	US-08-231-368-36	Sequence 36, Appl
31	11922	74.0	3011	1	US-08-440-210-36	Sequence 36, Appl
32	11922	74.0	3011	3	US-09-388-874-2	Sequence 2, Appli
33	11922	74.0	3011	3	US-09-046-604-36	Sequence 36, Appl
34	11922	74.0	3011	3	US-08-440-549-10	Sequence 10, Appl
35	11922	74.0	3011	4	US-09-916-359-2	Sequence 2, Appli
36	11915	73.9	3011	1	US-08-188-281B-1	Sequence 1, Appli
37	11915	73.9	3011	1	US-08-453-552-1	Sequence 1, Appli
38	11915	73.9	3011	2	US-08-710-637-1	Sequence 1, Appli
39	11915	73.9	3011	5	PCT-US93-00907-1	Sequence 1, Appli
40	11915	73.9	3011	5	PCT-US94-07280-1	Sequence 1, Appli
41	11915	73.9	3011	5	PCT-US95-01087-1	Sequence 1, Appli
42	11900	73.8	3011	3	US-08-811-566-20	Sequence 20, Appl
43	11900	73.8	3011	3	US-09-034-756-20	Sequence 20, Appl
44	11857	73.6	2995	3	US-08-444-818-138	Sequence 138, App
45	11828	73.4	3011	5	PCT-US91-02225-10	Sequence 10, Appl

ALIGNMENTS

```
RESULT 1
US-07-925-695-5
; Sequence 5, Application US/07925695
; Patent No. 5428145
;  GENERAL INFORMATION:
;    APPLICANT:  OKAMOTO, Hiroaki
;    APPLICANT:  NAKAMURA, Tetsuo
;    TITLE OF INVENTION:  NON-A, NON-B HEPATITIS VIRUS GENOME,
;    TITLE OF INVENTION:  POLYNUCLEOTIDES, POLYPEPTIDES, ANTIGEN, ANTIBODY AND
;    TITLE OF INVENTION:  DETECTION SYSTEMS
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Beveridge, DeGrandi, Weilacher & Young
;      STREET:  1850 M Street, N.W., Suite 800
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  US
;      ZIP:  20036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/925,695
;      FILING DATE:  19920807
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 287402/91
;      FILING DATE:  09-AUG-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 360441/91
;      FILING DATE:  05-DEC-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Weilacher, Robert G.
;      REGISTRATION NUMBER:  20,531
;      REFERENCE/DOCKET NUMBER:  06/87-48009
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 659-2811
;      TELEFAX:  (202) 659-1462
;      TELEX:  WUI 64470
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3033 amino acids
;      TYPE:  AMINO ACID
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  linear
US-07-925-695-5

  Query Match             98.1%;  Score 15812;  DB 1;  Length 3033;
```

Best Local Similarity 97.7%; Pred. No. 0;
Matches 2964; Conservative 31; Mismatches 38; Indels 0; Gaps 0;

```
Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |:|||
Db        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNTSRCWI 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
              |||||||||:||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PVSPNVAVQQPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIIGGAHWGVMFGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
              ||||||||||||||||||||:||||||| |     || :| :| ||||||:||||||
Db        361 YFSMQGAWAKVVVILLLAAGVDAQTHTVGGSTAHNARTLTGMFSLGARQKIQLINTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
              ||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db        421 HINRTALNCNDSLHTGFLASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
              |||||||||||||||||||| :|||||||||||||||||||||||||||||||||||||||
Db        481 PEDMRPYCWHYPPRQCGVVSASSVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
              ||||||| ||||||||||:||||||||||||| ||||||| ||||||||||||||||:|
Db        541 LNSTRPPQGSWFGCTWMNSTGYTKTCGAPPCRIRADFNASMDLLCPTDCFRKHPDTTYIK 600

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
              |||||||||||||||||||:||||||||||||||||||||||||||||||||||||||||
Db        661 RDRSQLSPLLHSTTEWAILPCTYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
              |||:||||||||||||||||||||||||||||:|||||||||||||||||||||||||||
Db        721 EWVVLLFLLLADARVCACLWMLILLGQAEAALEKLVVLHAASAASCNGFLYFVIFFVAAW 780

Qy        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
              |||||||||||||||||| ||||||||||||||||||||||||:||||||||||||||||
Db        781 YIKGRVVPLATYSLTGLWSFGLLLLALPQQAYAYDASVHGQIGAALLVLITLFTLTPGYK 840

Qy        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
              |||||||||||||||| ||||||||||||||||||||||||||||||| ||||||||||||
Db        841 TLLSRFLWWLCYLLTLAEAMVQEWAPPMQVRGGRDGIIWAVAIFCPGVVFDITKWLLAVL 900

Qy        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
              |||||||||||||||||||||||||| |||||||||| ||||||||||||||||||||||
Db        901 GPAYLLKGALTRVPYFVRAHALLRMCTMVRHLAGGRYVQMVLLALGRWTGTYIYDHLTPM 960

Qy        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
              |||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SDWAANGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020

Qy       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080
              ||||||||||||||||||||||||||||||||||||||||||||:||||||||||:|||
Db       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIEVLSTVTQSFLGTTISG 1080

Qy       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140

Qy       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200

Qy       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVQYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260

Qy       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320

Qy       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380

Qy       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440
              |||||||||||||||||||||||||||||||||||||:|||||||||||||||||||||
Db       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNAVAYYRGLDVSVIPTQGDVVVVA 1440

Qy       1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500
              |||||||:|||||||||||||||||||||||||||||||| ||||||||||||||||||
Db       1441 TDALMTGFTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQTVPQDAVSRSQRRGRTGRGRL 1500

Qy       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560
              ||||||||||||||||||||||||||||||||||||:|||||||||||||||||||||||
Db       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPAETTVRLRAYFNTPGLPVCQDHLE 1560

Qy       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620

Qy       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 WLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680

Qy       1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740
              ||||||||||:||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TGCVCIIGRLHVNQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740

Qy       1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 QQASKQAQDIQPAVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800

Qy       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860

Qy       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920

Qy       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCSGSW 1980

Qy       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040
              |||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||||
Db       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFISCQKGYKGVWAGTGIMTTRCPCGANIS 2040

Qy       2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100
              ||||||||||||||||||||||||||||||||||||||||||||:||||||||||||||||
Db       2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKIAIWRVAASEYAEVTQH 2100

Qy       2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db       2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPIPKPFFRDEVSFCVGLNSFVV 2160
```

```
Qy       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220
             |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       2161 GSQLPCDPEPDTDVLTSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220

Qy       2221 THGKAYDVDMVDANLFMGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYMLPKKR 2280
             |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       2221 THGKAYDVDMVDANLFMGGDVTRIESESKVVVLDSLDPMVEERSDLEPSIPSEYMLPKKR 2280

Qy       2281 FPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLSEDSI 2340
             |||||||||||||||||||||||||||||||||||||:||||||||||||||||||||| ||
Db       2281 FPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPKKTPTPPPRRRRTVGLSESSI 2340

Qy       2341 GDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTPPDELALSETGSISSMPPLEGELG 2400
              |||||||||||||||||||||||| ||||||:|||||||||||||||||||||||||| |
Db       2341 ADALQQLAIKSFGQPPPSGDSGLSTGADAADSGSRTPPDELALSETGSISSMPPLEGEPG 2400

Qy       2401 DPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460
             |||||||||| |||||||  ||| ||||||||||||||||||||||||||||||||||||
Db       2401 DPDLEPEQVELQPPPQGGVVTPGSGSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460

Qy       2461 EKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIKLAASKV 2520
             |||||||||||||||||||||:|||||||||||||||| ||::|||||||||||||||||
Db       2461 EKLPINPLSNSLLRYHNKVYCTTSKSASLRAKKVTFDRMQALDAHYDSVLKDIKLAASKV 2520

Qy       2521 TARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSETPIPTTI 2580
             ||||||:||||||||||||||||||||||||||||||||||||||||::|||||||||||
Db       2521 TARLLTLEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDTQTPIPTTI 2580

Qy       2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPA 2640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPA 2640

Qy       2641 QRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTE 2700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2641 QRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTE 2700

Qy       2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPTMLVCGD 2760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPTMLVCGD 2760

Qy       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820

Qy       2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDT 2880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDT 2880

Qy       2881 LDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2881 LDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2940

Qy       2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000

Qy       3001 DIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
             ||||||||||||||| ||||||||||||||||||
Db       3001 DIYHSVSRARPRLLLLGLLLLFVGVGLFLLPAR 3033
```

```
RESULT 2
US-07-925-695-8
; Sequence 8, Application US/07925695
; Patent No. 5428145
;  GENERAL INFORMATION:
;    APPLICANT:  OKAMOTO, Hiroaki
;    APPLICANT:  NAKAMURA, Tetsuo
;    TITLE OF INVENTION:  NON-A, NON-B HEPATITIS VIRUS GENOME,
;    TITLE OF INVENTION:  POLYNUCLEOTIDES, POLYPEPTIDES, ANTIGEN, ANTIBODY AND
;    TITLE OF INVENTION:  DETECTION SYSTEMS
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Beveridge, DeGrandi, Weilacher & Young
;      STREET:  1850 M Street, N.W., Suite 800
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  US
;      ZIP:  20036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/925,695
;      FILING DATE:  19920807
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 287402/91
;      FILING DATE:  09-AUG-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 360441/91
;      FILING DATE:  05-DEC-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Weilacher, Robert G.
;      REGISTRATION NUMBER:  20,531
;      REFERENCE/DOCKET NUMBER:  06/87-48009
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 659-2811
;      TELEFAX:  (202) 659-1462
;      TELEX:  WUI 64470
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3033 amino acids
;      TYPE:  AMINO ACID
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  linear
US-07-925-695-8

  Query Match             86.6%;  Score 13966;  DB 1;  Length 3033;
  Best Local Similarity   83.8%;  Pred. No. 0;
  Matches 2543;  Conservative  246;  Mismatches  244;  Indels    0;  Gaps    0;

Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              |||||||||||||||||||||||||||| ||||||||||||:||| ||||||||:|
Db         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGCGWAGWLLSPRGSRPTWGPTDPRHRSRNLG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
              :||||:||||||||||||||:||||||||||||||||:|:|||||||||||||||||
Db        121 RVIDTITCGFADLMGYIPVVGAPVGGVARALAHGVRVLEDGINYATGNLPGCSFSIFLLA 180

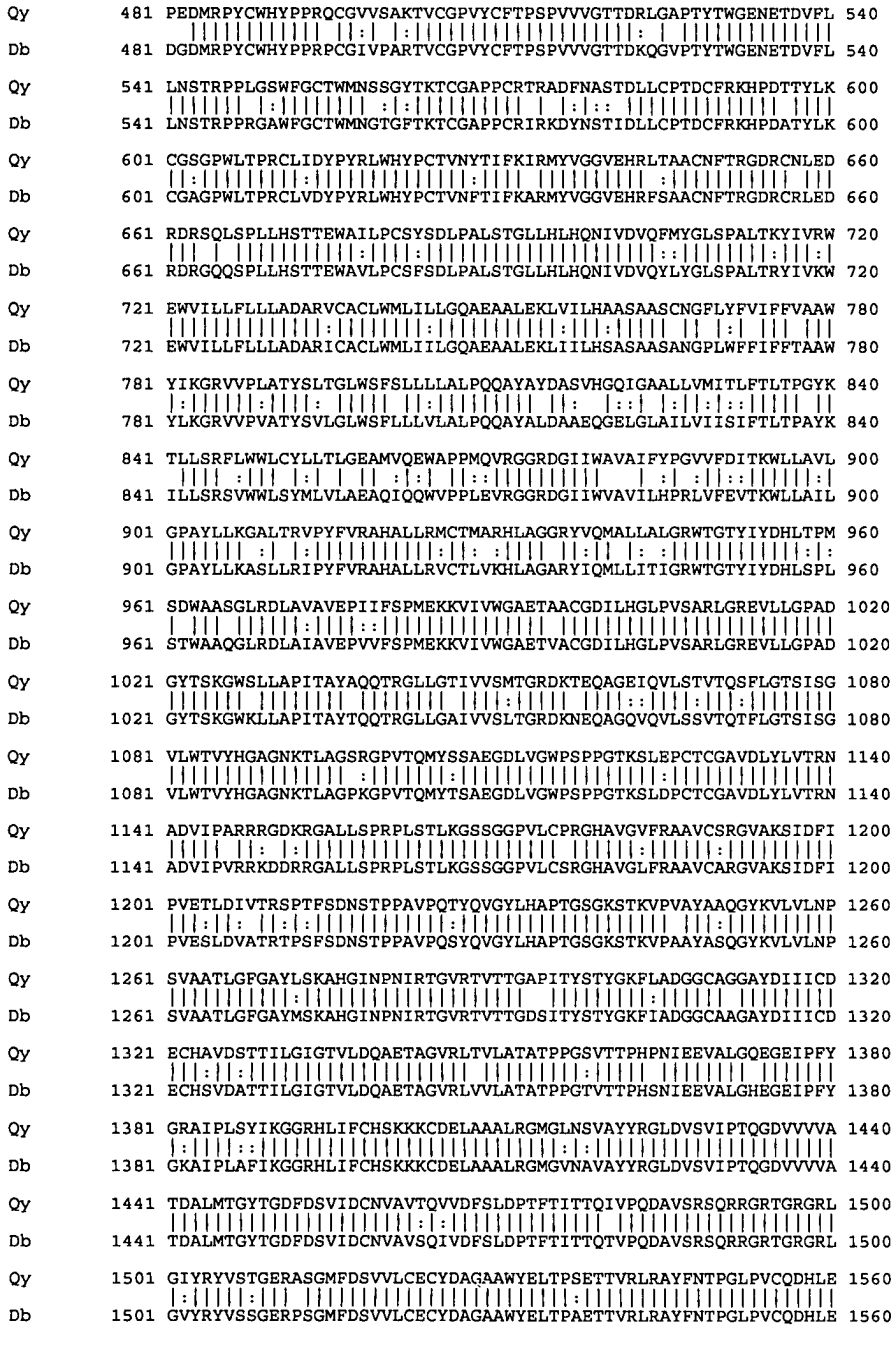
Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
              ||||:| |||| ||:|||: |  ||||:|:|||||  ||||:|||||||        |||
Db        181 LLSCVTVPVSAVEVRNISSSYYATNDCSNNSITWQLTDAVLHLPGCVPCENDNGTLHCWI 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
              |:|||||:  |||| |||||:||:||:|| ||||||||:|| ||: :| |:|||| | |
Db        241 QVTPNVAVKHRGALTRSLRTHVDMIVMAATACSALYVGDVCGAVMILSQAFMVSPQRHNF 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
              |:|||||| | |||||||||||::|||| ||||||| ||||:::||| | ||||:||||
Db        301 TQECNCSIYQGHITGHRMAWDMMLSWSPTLTMILAYAARVPELVLEIIFGGHWGVVFGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
              |||||||||||: |||| ||||| |:: |  | :|      ||  | :| : |:||||||
Db        361 YFSMQGAWAKVIAILLLVAGVDATTYSSGQEAGRTVAGFAGLFTTGAKQNLYLINTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
              |||||||||||| |||:|||||| |||||||||:|:|| :: ||:||| |:|| ||||
Db        421 HINRTALNCNDSLQTGFLASLFYTHKFNSSGCPERLSSCRGLDDFRIGWGTLEYETNVTN 480
```


Qy 481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
Db 481 DGDMRPYCWHYPPRPCGIVPARTVCGPVYCFTPSPVVVGTTDKQGVPTYTWGENETDVFL 540
Qy 541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
Db 541 LNSTRPPRGAWFGCTWMNGTGFTKTCGAPPCRIRKDYNSTIDLLCPTDCFRKHPDATYLK 600
Qy 601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
Db 601 CGAGPWLTPRCLVDYPYRLWHYPCTVNFTIFKARMYVGGVEHRFSAACNFTRGDRCRLED 660
Qy 661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
Db 661 RDRGQQSPLLHSTTEWAVLPCSFSDLPALSTGLLHLHQNIVDVQYLYGLSPALTRYIVKW 720
Qy 721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
Db 721 EWVILLFLLLADARICACLWMLIILGQAEAALEKLIILHSASAASANGPLWFFIFFTAAW 780
Qy 781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
Db 781 YLKGRVVPVATYSVLGLWSFLLLVLALPQQAYALDAAEQGELGLAILVIISIFTLTPAYK 840
Qy 841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
Db 841 ILLSRSVWWLSYMLVLAEAQIQQWVPPLEVRGGRDGIIWVAVILHPRLVFEVTKWLLAIL 900
Qy 901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
Db 901 GPAYLLKASLLRIPYFVRAHALLRVCTLVKHLAGARYIQMLLITIGRWTGTYIYDHLSPL 960
Qy 961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
Db 961 STWAAQGLRDLAIAVEPVVFSPMEKKVIVWGAETVACGDILHGLPVSARLGREVLLGPAD 1020
Qy 1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080
Db 1021 GYTSKGWKLLAPITAYTQQTRGLLGAIVVSLTGRDKNEQAGQVQVLSSVTQTFLGTSISG 1080
Qy 1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
Db 1081 VLWTVYHGAGNKTLAGPKGPVTQMYTSAEGDLVGWPSPPGTKSLDPCTCGAVDLYLVTRN 1140
Qy 1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200
Db 1141 ADVIPVRRKDDRRGALLSPRPLSTLKGSSGGPVLCSRGHAVGLFRAAVCARGVAKSIDFI 1200
Qy 1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260
Db 1201 PVESLDVATRTPSFSDNSTPPAVPQSYQVGYLHAPTGSGKSTKVPAAYASQGYKVLVLNP 1260
Qy 1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320
Db 1261 SVAATLGFGAYMSKAHGINPNIRTGVRTVTTGDSITYSTYGKFIADGGCAAGAYDIIICD 1320
Qy 1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380
Db 1321 ECHSVDATTILGIGTVLDQAETAGVRLVVLATATPPGTVTTPHSNIEEVALGHEGEIPFY 1380
Qy 1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440
Db 1381 GKAIPLAFIKGGRHLIFCHSKKKCDELAAALRGMGVNAVAYYRGLDVSVIPTQGDVVVVA 1440
Qy 1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500
Db 1441 TDALMTGYTGDFDSVIDCNVAVSQIVDFSLDPTFTITTQTVPQDAVSRSQRRGRTGRGRL 1500
Qy 1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560
Db 1501 GVYRYVSSGERPSGMFDSVVLCECYDAGAAWYELTPAETTVRLRAYFNTPGLPVCQDHLE 1560
Qy 1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
Db 1561 FWEAVFTGLTHIDAHFLSQTKQGGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
Qy 1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680
Db 1621 TLTGPTPLLYRLGAVTNEVTLTHPVTKYIATCMQADLEIMTSSWVLAGGVLAAVAAYCLA 1680
Qy 1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740
Db 1681 TGCISIIGRLHLNDRVVVAPDKEILYEAFDEMEECASKAALIEEGQRMAEMLKSKIQGLL 1740
Qy 1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
Db 1741 QQATRQAQDIQPAIQSSWPKLEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
Qy 1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860
Db 1801 LTSPLPTSTTILLNIMGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKILVDVLAGYG 1860
Qy 1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920
Db 1861 AGISGALVAFKIMSGEKPTVEDVVNLLPAILSPGALVVGVICAAILRRHVGQGEGAVQWM 1920
Qy 1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980
Db 1921 NRLIAFASRGNHVAPTHYVVESDASQRVTQVLSSLTITSLLRRLHAWITEDCPVPCSGSW 1980
Qy 1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040
Db 1981 LQDIWDWVCSILTDFKNWLSSKLLPKMPGIPFISCQKGYKGVWAGTGVMTTRCPCGANIS 2040
Qy 2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100
Db 2041 GHVRMGTMKITGPKTCLNLWQGTFPINCYTEGPCVPKPPPNYKTAIWRVAASEYVEVTQH 2100
Qy 2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160
Db 2101 GSFSYVTGLTSDNLKVPCQVPAPEFFSWVDGVQIHRFAPVPGPFFRDEVTFTVGLNSFVV 2160
Qy 2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220
Db 2161 GSQLPCDPEPDTEVLASMLTDPSHITAEAAARRLARGSPPSQASSSASQLSAPSLKATCT 2220
Qy 2221 THGKAYDVDMVDANLFMGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYMLPKKR 2280
Db 2221 THKTAYDCDMVDANLFMGGDVTRIESDSKVIVLDSLDSMTEVEDDREPSVPSEYLIKRRK 2280
Qy 2281 FPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLSEDSI 2340
Db 2281 FPPALPPWARPDYNPVLIETWKRPGYEPPTVLGCALPPTPQTPVPPPRRRRAKVLTQDNV 2340
Qy 2341 GDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTPPDELALSETGSISSMPPLEGELG 2400
Db 2341 EGVLREMADKVLSPLQDNNDSGHSTGADTGGDIVQQPSDETAASEAGSLSSMPPLEGEPG 2400
Qy 2401 DPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460
Db 2401 DPDLEFEPVGSAPPSEGECEVIDSDSKSWSTVSDQEDSVICCSMSYSWTGALITPCGPEE 2460
Qy 2461 EKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIKLAASKV 2520
Db 2461 EKLPINPLSNSLMRFHNKVYSTTSRSASLRAKKVTFDRVQVLDAHYDSVLQDVKRAASKV 2520
Qy 2521 TARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSETPIPTTI 2580
Db 2521 SARLLTVEEACALTPPHSAKSRYGFGAKEVRSLSRRAVNHIRSVWENLLEDQHTPIDTTI 2580
Qy 2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPA 2640
Db 2581 MAKNEVFCIDPTKGGKKPARLIVYPDLGVRVCEKMALYDIAQKLPKAIMGPSYGFQYSPA 2640
Qy 2641 QRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTE 2700

```
Db       2641 ERVDFLLKAWGSKKDPMGFSYDTRCFDSTVTERDIRTEESIYQACSLPQEARTVIHSLTE 2700

Qy       2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPTMLVCGD 2760

Db       2701 RLYVGGPMTNSKGQSCGYRRCRASGVFTTSMGNTMTCYIKALAACKAAGIVDPVMLVCGD 2760

Qy       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820

Db       2761 DLVVISESQGNEEDERNLRAFTEAMTRYSAPPGDLPRPEYDLELITSCSSNVSVALDSRG 2820

Qy       2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDT 2880

Db       2821 RRRYFLTRDPTTPITRAAWETVRHSPVNSWLGNIIQYAPTIWVRMVIMTHFFSILLAQDT 2880

Qy       2881 LDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2940

Db       2881 LNQNLNFEMYGAVYSVNPLDLPAIIERLHGLEAFSLHTYSPHELSRVAATLRKLGAPPLR 2940

Qy       2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000

Db       2941 AWKSRARAVRASLIAQGARAAICGRYLFNWAVKTKLKLTPLPEASRLDLSGWFTVGAGGG 3000

Qy       3001 DIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033

Db       3001 DIYHSVSHARPRLLLLCLLLLSVGVGIFLLPAR 3033


RESULT 3
US-07-925-695-9
; Sequence 9, Application US/07925695
; Patent No. 5428145
;  GENERAL INFORMATION:
;    APPLICANT:  OKAMOTO, Hiroaki
;    APPLICANT:  NAKAMURA, Tetsuo
;    TITLE OF INVENTION:  NON-A, NON-B HEPATITIS VIRUS GENOME,
;    TITLE OF INVENTION:  POLYNUCLEOTIDES, POLYPEPTIDES, ANTIGEN, ANTIBODY AND
;    TITLE OF INVENTION:  DETECTION SYSTEMS
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Beveridge, DeGrandi, Weilacher & Young
;      STREET:  1850 M Street, N.W., Suite 800
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  US
;      ZIP:  20036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/925,695
;      FILING DATE:  19920807
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 287402/91
;      FILING DATE:  09-AUG-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 360441/91
;      FILING DATE:  05-DEC-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Weilacher, Robert G.
;      REGISTRATION NUMBER:  20,531
;      REFERENCE/DOCKET NUMBER:  06/87-48009
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 659-2811
;      TELEFAX:  (202) 659-1462
;      TELEX:  WUI 64470
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3033 amino acids
;      TYPE:  AMINO ACID
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  linear
US-07-925-695-9

  Query Match           86.1%;  Score 13881;  DB 1;  Length 3033;
  Best Local Similarity  83.5%;  Pred. No. 0;
  Matches 2534;  Conservative  245;  Mismatches  254;  Indels    0;  Gaps    0;

Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120

Db         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGCGWAGWLLSPRGSRPTWGPTDPRHRSRNLG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180

Db        121 RVIDTITCGFADLMGYIPVVGAPVGGVARALAHGVRVLEDGINYATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240

Db        181 LLSCVTMPVSAVEVRNISSSYYATNDCSNNSITWQLTDAVLHLPGCVPCENDNGTLRCWI 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300

Db        241 QVTPDVAVKHRGALTRSLRTHVDMIVMAATACSALYVGDVCGAVMILSQAFMVSPQRHNF 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360

Db        301 TQECNCSIYQGHITGHRMAWDMMLNWSPTLAMILAYAARVPELVLEIIFGGHWGVAFGLG 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420

Db        361 YFSMQGAWAKVVAILLLVAGVDASTYSTGQQAGRAAYGISSLFNTGAKQNLHLINTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480

Db        421 HINRTALNCNDSLETGFIASLVYYRRFNSSGCPERLSSCRGLDDPRIGWGTLEYETNVTN 480

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540

Db        481 DEDMRPYCWHYPPRPCGIVPARTVCGPVYCFTPSPVVVGTTDKQGVPTYTWGENETDVFL 540

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600

Db        541 LNSTRPPRGAWFGCTWMNGTGFTKTCGAPPCRIRKDYNSTIDLLCPTDCFRKHPDATYLK 600

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660

Db        601 CGAGPWLTPRCLVDYPYRLWHYPCTVNFTIFKARMYVGGVEHRFSAACNFTRGDRCRLED 660

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720

Db        661 RDRGQQSPLLHSTTEWAVFPCSFSDLPALSTGLLHLHQNIVDVQYLYGLSPALTRYIVKW 720

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780

Db        721 EWVILLFLLLADARVCACLWMLNILGQAEAALEKLIILHSASAASANGPLWFFIFFTAAW 780

Qy        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840

Db        781 YLKGRVVPVATYSVLGLWSFLLLVLALPQQAYALDAAEQGELGLAILVIISIFTLTPAYK 840

Qy        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900

Db        841 ILLSRSVWWLSYMLVLAEAQIQQWVPPLEVRGGRDGIIWVAVILHPRLVFEVTKWLLAIL 900

Qy        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960

Db        901 GPAYLLRASLLRIPYFVRAHALLRVCTLVKHLAGARYIQMLLITIGRWTGTYIYDHLSPL 960

Qy        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
```

```
Db        961 STWAAQGLRDLAIAVEPVVFSPMEKKVIVWGAETVACGDILHGLPVSARLGREVLLGPAD 1020

Qy       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080

Db       1021 GYTSKGWNLLAPITAYTQQTRGLLGAIVVSLTGRDKNEQAGQVQVLSSVTQTFLGTSISG 1080

Qy       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140

Db       1081 VLWTVYHGAGNKTLAGPKGPVTQMYTSAEGDLVGWPSPPGTKSLDPCTCGAVDLYLVTRN 1140

Qy       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200

Db       1141 ADVIPVRRKDDRRGALLSPRPLSTLKGSSGGPVLCSRGHAVGLFRAAVCARGVAKSIDFI 1200

Qy       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260

Db       1201 PVESLDIATRTPSFSDNSAPPAVPQSYQVGYLHAPTGSGKSTKVPAAYASQGYKVLVLNP 1260

Qy       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320

Db       1261 SVAATLGFGAYMSKAHGINPNIRTGVRTVTTGDSITYSTYGKFIADGGCAAGAYDIIICD 1320

Qy       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380

Db       1321 ECHSVDATTILGIGTVLDQAETAGVRLVVLATATPPGTVTTPHSNIEEVALGHEGEIPFY 1380

Qy       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440

Db       1381 GKAIPLAFIKGGRHLIFCHSKKKCDELAAALRGTGVNAVAYYRGLDVSVIPTQGDVVVVA 1440

Qy       1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500

Db       1441 TDALMTGYTGDFDSVIDCNVAVSQIVDFSLDPTFTITTQTVPQDAVSRSQRRGRTGRGRL 1500

Qy       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560

Db       1501 GIYRYVSSGEGPSGMFDSVVPCECYDAGAAWYELTPAETTVRLRAYFNTPGLPVCQDHLE 1560

Qy       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620

Db       1561 FWEAVFTGLTHINAHFLSQTKQGGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620

Qy       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680

Db       1621 TLTGPTPLLYRLGAVTNEVTLTHPVTKYIATCMQADLEIMTSSWVLAGGVLAAVAAYCLA 1680

Qy       1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740

Db       1681 TGCISIIGRLHLNDRVVVTPDKEILYEAFDEMEECASKAALIEEGQRMAEMLKSKIQGLL 1740

Qy       1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800

Db       1741 QQATRQAQGMQPAIQSSWPKLEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800

Qy       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860

Db       1801 LTSPLPTSTTILLNIMGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKILVDVLAGYG 1860

Qy       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920

Db       1861 AGISGALVAFKIMSGEKPTVEDVVNLLPAILSPGALVVGVICAAILRRHVGQGEGAVQWM 1920

Qy       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980

Db       1921 NRLIAFASRGNHVAPTHYVVESDASQRVTQVLSSLTITSLLRRLHAWITEDCPIPCSGSW 1980

Qy       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040

Db       1981 LQDIWDWVCSILTDFKNWLSSKLLPKMPGIPFISCQKGYKGVWAGTGVMTTRYPCGANIS 2040

Qy       2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100

Db       2041 GHVRMGTMKITGPKTCLNLWQGTFPINCYTEGPCVPKPPPNYKTAIWRVAASEYVEVTQH 2100

Qy       2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160

Db       2101 GSFSYVTGLTSDNLKVPCQVPAPEFFSWVDGVQIHRFAPVPGPFFRDEVTFTVGLNSFVV 2160

Qy       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220

Db       2161 GSQLPCDPEPDTEVLASMLTDPSHITAEAAARRLARGSPPSQASSSASQLSAPSLKATCT 2220

Qy       2221 THGKAYDVDMVDANLFMGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYMLPKKR 2280

Db       2221 THKTAYDCDMVDANLFMGGDVTRIESDSKVIVLDSLDSMTEVEDDREPSVPSEYLIKRRK 2280

Qy       2281 FPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLSEDSI 2340

Db       2281 FPPALPPWARPDYNPVLIETWKRPGYEPPTVLGCALPPTLQTPVPPPRRRRAKILTQDDV 2340

Qy       2341 GDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTPPDELALSETGSISSMPPLEGELG 2400

Db       2341 EGILREMADKVLSPLQDNNDSGHSTGADTGGDIVQQPSDETAASEAGSLSSMPPLEGEPG 2400

Qy       2401 DPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEEDDSVVCCSMSYSWTGALITPCSPEE 2460

Db       2401 DPDLEFEPVGSAPPSEGECEVIDSDSKSWSTVSDQEDSVICCSMSYSWTGALITPCGPEE 2460

Qy       2461 EKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIKLAASKV 2520

Db       2461 EKLPINPLSNSLMRFHNKVYSTTSRSASLRAKKVTFDRVQVLDAHYDSVLQDVKRAASKV 2520

Qy       2521 TARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSETPIPTTI 2580

Db       2521 GARLLTVEEACALTPPHSAKSRYGFGAKEVRSLSRRAVNHIRSVWENLLEDQRTPIDTTI 2580

Qy       2581 MAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYGFQYSPA 2640

Db       2581 MAKNEVFCIDPTKGGKKPARLIVYPDLGVRVCEKMALYDITQKLPKAIMGPSYGFQYSPA 2640

Qy       2641 QRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTAIHSLTE 2700

Db       2641 ERVDFLLKAWGSKKDPMGFSYDTRCFDSTVTERDIRTEESIYQACSLPQEARTVIHSLTE 2700

Qy       2701 RLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPTMLVCGD 2760

Db       2701 RLYVGGPMTNSKGQSCGYRRCRASGVFTTSMGNTMTCYIKALAACKAAGIVDPVMLVCGD 2760

Qy       2761 DLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSVALGPQG 2820

Db       2761 DLVVISESQGNEEDERNLRAFTEAMTRYSAPPGDLPRPEYDLELITSCSSNVSVALDSRG 2820

Qy       2821 RRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSILMAQDT 2880

Db       2821 RRRYFLTRDPTTPITRAAWETVRHSPVNSWLGNIIQYAPTIWVRMVIMTHFFSILLAQDT 2880

Qy       2881 LDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKLGAPPLR 2940

Db       2881 LNQNLNFEMYGAVYSVNPLDLPAIIERLHGLEAFSLHTYSPHELSRVAATLRKLGAPPLR 2940

Qy       2941 AWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFTVGAGGG 3000

Db       2941 AWKSRARAVRASLIAQGARAAICGRYLFNWAVKTKLKLTPLPEASRLDLSGWFTVGAGGG 3000

Qy       3001 DIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033

Db       3001 DIYHSVSHARPRLLLLCLLLLSVGVGIFLLPAR 3033
```

RESULT 4
US-09-539-601-3
; Sequence 3, Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
; APPLICANT: Bartenschlager, Ralf FW

; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
; FILE REFERENCE: all sequences
; CURRENT APPLICATION NUMBER: US/09/539,601C
; CURRENT FILING DATE: 2001-08-30
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
; EARLIER FILING DATE: 1999-04-03
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 3010
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-539-601-3

Query Match 74.6%; Score 12027.5; DB 4; Length 3010;
Best Local Similarity 71.8%; Pred. No. 0;
Matches 2182; Conservative 352; Mismatches 470; Indels 35; Gaps 7;

```
Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              ||||||| |:  |::| :|||||||||||||||||||||||||||||| ||| ||||:|
Db         61 RRQPIPKARQPEGRAWAQPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPTDPRRRSRNLG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
              |||||||||||||||||:|||||||  ||||||||||||||||:||||||||||||||||
Db        121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
              ||||:| | || ||:|:|  | |||||:| || ::    ::| |||||| :  |:|:||:
Db        181 LLSCLTIPASAYEVRNVSGVYHVTNDCSNASIVYEAADMIMHTPGCVPCVRENNSSRCWV 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
              ::| :| :     |  :| |:|::| :| ||||:|||||| | | ||:|  ||: |
Db        241 ALTPTLAARNASVPTTTIRRHVDLLVGAAALCSAMYVGDLCGSVFLVAQLFTFSPRRHET 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
              |||||||||||  :|||||||||||||||||  ::::  :|:|: ::|:::||||||: |||
Db        301 VQDCNCSIYPGHVTGHRMAWDMMMNWSPTAALVVSQLLRIPQAVVDMVAGAHWGVLAGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
              |:|| | ||||::::|| ||||  |:  ||: |: |  :|||  |  ||||||||||||
Db        361 YYSMVGNWAKVLIVMLLFAGVDGGTYVTGGTMAKNTLGITSLFSPGSSQKIQLVNTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
              |||||||||||||:|||:|:||| | ||||||||::|  |:||  ||| : |  | ::
Db        421 HINRTALNCNDSLNTGFLAALFYVHKFNSSGCPERMASCSPIDAFAQGWGPITY--NESH 478

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
              | |||||| || ||:| |  |||||||||||||||||||| | |||:|||||||| |
Db        479 SSDQRPYCWHYAPRPCGIVPAAQVCGPVYCFTPSPVVVGTTDRFGVPTYSWGENETDVLL 538

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
              ||:|||| |:|||||||||:|:|||||  |||       | :  | ||||||||||: || |
Db        539 LNNTRPPQGNWFGCTWMNSTGFTKTCGGPPCNIGGIGNKT--LTCPTDCFRKHPEATYTK 596

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
              ||||||||||:  |||||||||||||:|||:|||||||||| ||||:|||:||||||
Db        597 CGSGPWLTPRCLVHYPYRLWHYPCTVNFTIFKVRMYVGGVEHRLEAACNWTRGERCNLED 656

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
              ||||:||||| ||||| :||||:: |||||||:|||||:||||::||:  |:  : ::|
Db        657 RDRSELSPLLLSTTEWQVLPCSFTTLPALSTGLIHLHQNVVDVQYLYGIGSAVVSPAIKW 716

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
              |:|:|||||||||||||::|: |||||||  ||:|:||| |  :| | |::|| |||
Db        717 EYVLLLFLLLADARVCACLWMMLLIAQAEAALENLVVLNAASVAGAHGILSFLVFFCAAW 776

Qy        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
              |||||:|| | |:| |:|   ||||||| :||| |  :    | |: | : | ||:| ||
Db        777 YIKGRLVPGAAYALYGVWPLLLLLLALPPRAYAMDREMAASCGGAVFVGLILLTLSPHYK 836

Qy        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
              |:| :||| | :|  || :| | ||: |||||| :|      :| ::| ||| |||:|
Db        837 LFLARLIWWLQYFITRAEAHLQVWIPPLNVRGGRDAVILLTCAIHPELIFTITKILLAIL 896

Qy        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
              ||  :|:  :|:|||||||| |:| | : | :||| ||||||: |   ||||:||||||:
Db        897 GPLMVLQAGITKVPYFVRAHGLIRACMLVRKVAGGHYVQMALMKLAALTGTYVYDHLTPL 956

Qy        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
              ||| :|||||||||||::|| || ||| |||:|||||||: |||||||  |||: |||||
Db        957 RDWAHAGLRDLAVAVEPVVFSDMETKVITWGADTAACGDIILGLPVSARRGREIHLGPAD 1016

Qy       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080
                :|| ||||||||:|||||||| |: |:||||: :  ||:||:|| ||||| | ::|
Db       1017 SLEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDRNQVEGEVQVVSTATQSFLATCVNG 1076

Qy       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
              | |||||||:||||| :||:||||:: : ||||| :||| :|| |||||: |||||||:
Db       1077 VCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRH 1136

Qy       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200
              ||||| ||||| ||:|||||:| ||||||||:||| |||||:|||||:|||||::||:
Db       1137 ADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFV 1196

Qy       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260
              |||:::   ||| |:|||:|||||||:|| :|||||||||||| ||||||||||||||
Db       1197 PVESMETTMRSPVFTDNSSPPAVPQTFQVAHLHAPTGSGKSTKVPAAYAAQGYKVLVLNP 1256

Qy       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320
              ||||||||||:||||||:|||||||||:|||||||||||||||||||:||||||||||
Db       1257 SVAATLGFGAYMSKAHGIDPNIRTGVRTITTGAPITYSTYGKFLADGGCSGGAYDIIICD 1316

Qy       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380
              |||: |||||||||||||||||| || |||||||||| ||||||||||    ||||||
Db       1317 ECHSTDSTTILGIGTVLDQAETAGARLVVLATATPPGSVTVPHPNIEEVALSSTGEIPFY 1376

Qy       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440
              |:|||:  |||||||||||||||||||| | |:|||:|||||||||||| |||:|||
Db       1377 GKAIPIETIKGGRHLIFCHSKKKCDELAAKLSGLGLNAVAYYRGLDVSVIPTSGDVIVVA 1436

Qy       1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500
              |||||||:|||||||||||  ||| |||||||||||| |  |||||||||||||||||:
Db       1437 TDALMTGFTGDFDSVIDCNTCVTQTVDFSLDPTFTIETTTVPQDAVSRSQRRGRTGRGRM 1496

Qy       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560
              ||||:|:|||| |||||| |||||||| |||||||:||:|||||| ||||||||||||
Db       1497 GIYRFVTPGERPSGMFDSSVLCECYDAGCAWYELTPAETSVRLRAYLNTPGLPVCQDHLE 1556

Qy       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
              |||:|||||||||||||:|:|| || |||||||||:|||||| ||||| ||||
Db       1557 FWESVFTGLTHIDAHFLSQTKQAGDNFPYLVAYQATVCARAQAPPPSWDQMWKCLIRLKP 1616

Qy       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680
              || |||||||||:| |||| |||:|||| || |||||:|||||| ||||||:|||||
Db       1617 TLHGPTPLLYRLGAVQNEVTTTHPITKYIMACMSADLEVVTSTWVLVGGVLAALAAYCLT 1676

Qy       1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740
              || | |:||: :: :  : ||:|||| |||||||||     ||:| ::||  | |  |||
Db       1677 TGSVVIVGRIILSGKPAIIPDREVLYREFDEMEECASHLPYIEQGMQLAEQFKQKAIGLL 1736

Qy       1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
              | |:|||:  | |:: |  :| ||||||||||||||||||||||||:||:|||:|:
Db       1737 QTATKQAEAAAPVVESKWRTLEAFWAKHMWNFISGIQYLAGLSTLPGNPAIASLMAFTAS 1796

Qy       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860
              :||||:|  |:| ||||||:|:|:|||: |: || :|: |||||||||||||||||||
Db       1797 ITSPLTTQHTLLFNILGGWVAAQLAPPSAASAFVGAGIAGAAVGSIGLGKVLVDILAGYG 1856

Qy       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920
              ||::|||||||:|||| || ||:|||| ||||||||||:||||||||||||||||||||
Db       1857 AGVAGALVAFKVMSGEMPSTEDLVNLLPAILSPGALVVGVVCAAILRRHVGPGEGAVQWM 1916
```

```
Qy       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980
Db       1917 NRLIAFASRGNHVSPTHYVPESDAAARVTQILSSLTITQLLKRLHQWINEDCSTPCSGSW 1976

Qy       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040
Db       1977 LRDVWDWICTVLTDFKTWLQSKLLPRLPGVPFFSCQRGYKGVWRGDGIMQTTCPCGAQIT 2036

Qy       2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100
Db       2037 GHVKNGSMRIVGPRTCSNTWHGTFPINAYTTGPCTPSPAPNYSRALWRVAAEEYVEVTRV 2096

Qy       2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160
Db       2097 GDFHYVTGMTTDNVKCPCQVPAPEFFTEVDGVRLHRYAPACKPLLREEVTFLVGLNQYLV 2156

Qy       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220
Db       2157 GSQLPCEPEPDVAVLTSMLTDPSHITAETAKRRLARGSPPSLASSSASQLSAPSLKATCT 2216

Qy       2221 THGKAYDVDMVDANLF----MGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYML 2276
Db       2217 TRHDSPDADLIEANLLWRQEMGGNITRVESENKVVILDSFEPLQAEEDEREVSVPAEILR 2276

Qy       2277 PKKRFPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLS 2336
Db       2277 RSRKFPRAMPIWARPDYNPPLLESWKDPDYVPPVVHGCPLPPAKAPPIPPPRRKRTVVLS 2336

Qy       2337 EDSIGDALQQLAIKSFGQPPPSG-DSGLSTGAGAADSGSQTPPDELALSETGSISSMPPL 2395
Db       2337 ESTVSSALAELATKTFGSSESSAVDSGTATA-----SPDQPSDDGDAGSDVESYSSMPPL 2391

Qy       2396 EGELGDPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEE-DDSVVCCSMSYSWTGALIT 2454
Db       2392 EGEPGDPDL--------------------SDGSWSTVSEEASEDVVCCSMSYTWTGALIT 2431

Qy       2455 PCSPEEEKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIK 2514
Db       2432 PCAAEETKLPINALSNSLLRHHNLVYATTSRSASLRQKKVTFDRLQVLDDHYRDVLKEMK 2491

Qy       2515 LAASKVTARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSET 2574
Db       2492 AKASTVKAKLLSVEEACKLTPPHSARSKFGYGAKDVRNLSSKAVNHIRSVWKDLLEDTET 2551

Qy       2575 PIPTTIMAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYG 2634
Db       2552 PIDTTIMAKNEVFCVQPEKGGRKPARLIVFPDLGVRVCEKMALYDVVSTLPQAVMGSSYG 2611

Qy       2635 FQYSPAQRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTA 2694
Db       2612 FQYSPGQRVEFLVNAWKAKKCPMGFAYDTRCFDSTVTENDIRVEESIYQCCDLAPEARQA 2671

Qy       2695 IHSLTERLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPT 2754
Db       2672 IRSLTERLYIGGPLTNSKGQNCGYRRCRASGVLTTSCGNTLTCYLKAAAACRAAKLQDCT 2731

Qy       2755 MLVCGDDLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSV 2814
Db       2732 MLVCGDDLVVICESAGTQEDEASLRAFTEAMTRYSAPPGDPPKPEYDLELITSCSSNVSV 2791

Qy       2815 ALGPQGRRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSI 2874
Db       2792 AHDASGKRVYYLTRDPTTPLARAAWETARHTPVNSWLGNIIMYAPTLWARMILMTHFFSI 2851

Qy       2875 LMAQDTLDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKL 2934
Db       2852 LLAQEQLEKALDCQIYGACYSIEPLDLPQIIQRLHGLSAFSLHSYSPGEINRVASCLRKL 2911

Qy       2935 GAPPLRAWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFT 2994
Db       2912 GVPPLRVWRHRARSVRARLLSQGGRAATCGKYLFNWAVRTKLKLTPIPAASQLDLSSWFV 2971

Qy       2995 VGAGGGDIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
Db       2972 AGYSGGDIYHSLSRARPRWFMWCLLLLSVGVGIYLLPNR 3010
```

RESULT 5
US-09-539-601-21
; Sequence 21, Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
; APPLICANT: Bartenschlager, Ralf FW
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
; FILE REFERENCE: all sequences
; CURRENT APPLICATION NUMBER: US/09/539,601C
; CURRENT FILING DATE: 2001-08-30
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
; EARLIER FILING DATE: 1999-04-03
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 3010
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-539-601-21

Query Match 74.5%; Score 12010.5; DB 4; Length 3010;
Best Local Similarity 71.7%; Pred. No. 0;
Matches 2178; Conservative 355; Mismatches 471; Indels 35; Gaps 7;

```
Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
Db          1 MGTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
Db         61 RRQPIPKARQPEGRAWAQPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPTDPRRRSRNLG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
Db        121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
Db        181 LLSCLTIPASAYEVRNVSGVYHVTNDCSNASIVYEAADMIMHTPGCVPCVRENNSSRCWV 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
Db        241 ALTPTLAARNASVPTTTIRRHVDLLVGAAALCSAMYVGDLCGSVFLVAQLFTFSPRRHET 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
Db        301 VQDCNCSIYPGHVTGHRMAWDMMMNWSPTAALVVSQLLRIPQAVVDMVAGAHWGVLAGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
Db        361 YYSMVGNWAKVLIVMLLFAGVDGGTYVTGGTMAKNTLGITSLFSPGSSQKIQLVNTNGSW 420

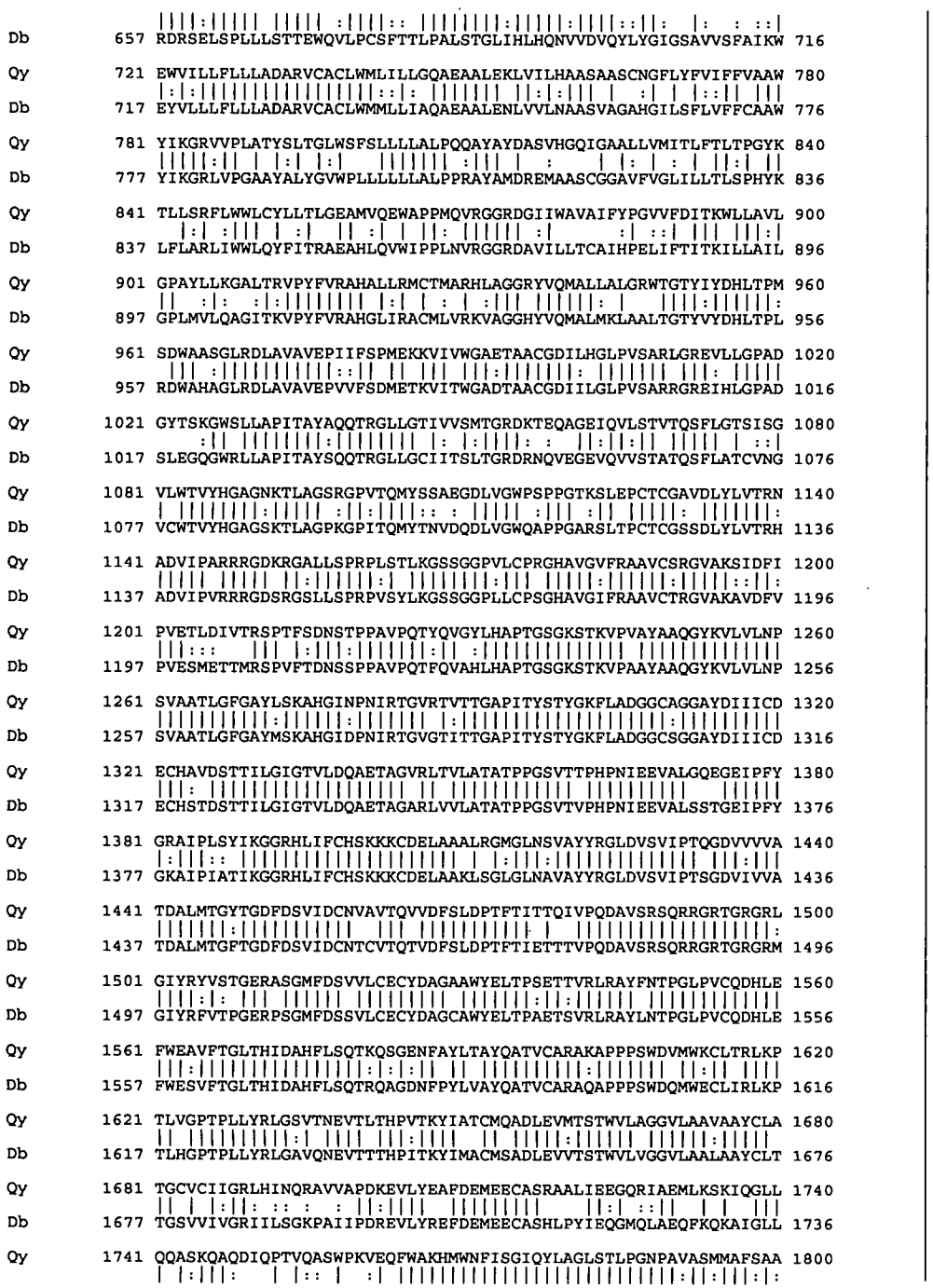
Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
Db        421 HINRTALNCNDSLNTGFLAALFYVHKFNSSGCPERMASCSPIDAFAQGWGPITY--NESH 478

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
Db        479 SSDQRPYCWHYAPRPCGIVPAAQVCGPVYCFTPSPVVVGTTDRFGVPTYSWGENETDVLL 538

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
Db        539 LNNTRPPQGNWFGCTWMNSTGFTKTCGGPPCNIGGIGNKT--LTCPTDCFRKHPEATYTK 596

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
Db        597 CGSGPWLTPRCLVHYPYRLWHYPCTVNFTIFKVRMYVGGVEHRLEAACNWTRGERCNLED 656

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
```


Db 657 RDRSELSPLLLSTTEWQVLPCSFTTLPALSTGLIHLHQNVVDVQYLYGIGSAVVSFAIKW 716
Qy 721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
Db 717 EYVLLLFLLLADARVCACLWMMLLIAQAEAALENLVVLNAASVAGAHGILSFLVFFCAAW 776
Qy 781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
Db 777 YIKGRLVPGAAYALYGVWPLLLLLLALPPRAYAMDREMAASCGGAVFVGLILLTLSPHYK 836
Qy 841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
Db 837 LFLARLIWWLQYFITRAEAHLQVWIPPLNVRGGRDAVILLTCAIHPELIFTITKILLAIL 896
Qy 901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
Db 897 GPLMVLQAGITKVPYFVRAHGLIRACMLVRKVAGGHYVQMALMKLAALTGTYVYDHLTPL 956
Qy 961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
Db 957 RDWAHAGLRDLAVAVEPVVFSDMETKVITWGADTAACGDIILGLPVSARRGREIHLGPAD 1016
Qy 1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080
Db 1017 SLEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDRNQVEGEVQVVSTATQSFLATCVNG 1076
Qy 1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
Db 1077 VCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRH 1136
Qy 1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200
Db 1137 ADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFV 1196
Qy 1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260
Db 1197 PVESMETTMRSPVFTDNSSPPAVPQTFQVAHLHAPTGSGKSTKVPAAYAAQGYKVLVLNP 1256
Qy 1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320
Db 1257 SVAATLGFGAYMSKAHGIDPNIRTGVGTITTGAPITYSTYGKFLADGGCSGGAYDIIICD 1316
Qy 1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380
Db 1317 ECHSTDSTTILGIGTVLDQAETAGARLVVLATATPPGSVTVPHPNIEEVALSSTGEIPFY 1376
Qy 1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440
Db 1377 GKAIPIATIKGGRHLIFCHSKKKCDELAAKLSGLGLNAVAYYRGLDVSVIPTSGDVIVVA 1436
Qy 1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500
Db 1437 TDALMTGFTGDFDSVIDCNTCVTQTVDFSLDPTFTIETTTVPQDAVSRSQRRGRTGRGRM 1496
Qy 1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560
Db 1497 GIYRFVTPGERPSGMFDSSVLCECYDAGCAWYELTPAETSVRLRAYLNTPGLPVCQDHLE 1556
Qy 1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
Db 1557 FWESVFTGLTHIDAHFLSQTRQAGDNFPYLVAYQATVCARAQAPPPSWDQMWECLIRLKP 1616
Qy 1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680
Db 1617 TLHGPTPLLYRLGAVQNEVTTTHPITKYIMACMSADLEVVTSTWVLVGGVLAALAAYCLT 1676
Qy 1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740
Db 1677 TGSVVIVGRIILSGKPAIIPDREVLYREFDEMEECASHLPYIEQGMQLAEQFKQKAIGLL 1736
Qy 1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
Db 1737 QTATKQAEAAAPVVESKWRTLEAFWAKHMWNFISGIQYLAGLSTLPGNPAIASLMAFTAS 1796
Qy 1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860
Db 1797 ITSPLTTQHTLLFNILGGWVAAQLAPPSAASAFVGAGIAGAAVGSIGLGKVLVDILAGYG 1856
Qy 1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920
Db 1857 AGVAGALVAFKVMSGEMPSTEDLVNLLPAILSPGALVVGVVCAAILRRHVGPGEGAVQWM 1916
Qy 1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980
Db 1917 NRLIAFASRGNHVSPTHYVSESDAAARVTQILSSLTITQLLKRLHQWINEDCSTPCSGSW 1976
Qy 1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040
Db 1977 LRDVWDWICTVLTDFKTWLQSKLLPRLPGVPFFSCQRGYKGVWRGDGIMQTTCPCGAQIT 2036
Qy 2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100
Db 2037 GHVKNGSMRIVGPRTCSNTWHGTFPINAYTTGPCTPSPAPNYSRALWRVAAEEYVEVTRV 2096
Qy 2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160
Db 2097 GDFHYVTGMTTDNVKCPCQVPAPEFFTEVDGVRLHRYAPACKPLLREEVTFLVGLNQYLV 2156
Qy 2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220
Db 2157 GSQLPCGPEPDVAVLTSMLTDPSHITAETAKRRLARGSPPSLASSSASQLSAPSLKATCT 2216
Qy 2221 THGKAYDVDMVDANLF----MGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYML 2276
Db 2217 TRHDSPDADLIEANLLWRQEMGGNITRVESENKVVILDSFEPLQAEEDEREVSVPAEILR 2276
Qy 2277 PKKRFPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLS 2336
Db 2277 RSRKFPRAMPIWARPDYNPPLLESWKDPDYVPPVVHGCPLPPAKAPPIPPPRRERTVVLS 2336
Qy 2337 EDSIGDALQQLAIKSFGQPPPSG-DSGLSTGAGAADSGSQTPPDELALSETGSISSMPPL 2395
Db 2337 ESTVSSALAELATKTFGSSESSAVDSGTATA-----SPDQPSDDGDAGSDVESYSSMPPL 2391
Qy 2396 EGELGDPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEE-DDSVVCCSMSYSWTGALIT 2454
Db 2392 EGEPGDPDL----------SDGSWSTVSEEASEDVVCCSMSYTWTGALIT 2431
Qy 2455 PCSPEEEKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIK 2514
Db 2432 PCAAEETKLPVNALSNSLLRHHNLVYATTSRSASLRQKKVTFDRLQVLDDHYRDVLKEMK 2491
Qy 2515 LAASKVTARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSET 2574
Db 2492 AKASTVKAKLLSVEEACKLTPPHSARSKFGYGAKDVRNLSSKAVNHIRSVWKDLLEDTET 2551
Qy 2575 PIPTTIMAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYG 2634
Db 2552 PIDTTIMAKNEVFCVQPEKGGRKPARLIVFPDLGVRVCEKMALYDVVSTLPQAVMGSSYG 2611
Qy 2635 FQYSPAQRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTA 2694
Db 2612 FQYSPGQRVEFLVNAWKAKKCPMGFAYDTRCFDSTVTENDIRVEESIYQCCDLAPEARQA 2671
Qy 2695 IHSLTERLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPT 2754
Db 2672 IRSLTERLYIGGPLTNSKGQNCGYRRCRASGVLTTSCGNTLTCYLKAAAACRAAKLQDCT 2731
Qy 2755 MLVCGDDLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSV 2814
Db 2732 MLVCGDDLVVICESAGTQEDEASLRAFTEAMTRYSAPPGDPPKPEYDLELITSCSSNVSV 2791
Qy 2815 ALGPQGRRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSI 2874
Db 2792 AHDASGKRVYYLTRDPTTPLARAAWETARHTPVNSWLGNIIMYAPTLWARMILMTHFFSI 2851

```
Qy       2875 LMAQDTLDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKL 2934
Db       2852 LLAQEQLEKALDCQIYGACYSIEPLDLPQIIQRLHGLSAFSLHSYSPGEINRVASCLRKL 2911

Qy       2935 GAPPLRAWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFT 2994
Db       2912 GVPPLRVWRHRARSVRARLLSQGGRAATCGKYLFNWAVRTKLKLTPIPAASQLDLSSWFV 2971

Qy       2995 VGAGGGDIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
Db       2972 AGYSGGDIYHSLSRARPRWFMWCLLLLSVGVGIYLLPNR 3010


RESULT 6
US-09-539-601-27
; Sequence 27, Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
;  APPLICANT: Bartenschlager, Ralf FW
;  TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
;  FILE REFERENCE: all sequences
;  CURRENT APPLICATION NUMBER: US/09/539,601C
;  CURRENT FILING DATE: 2001-08-30
;  EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
;  EARLIER FILING DATE: 1999-04-03
;  NUMBER OF SEQ ID NOS: 51
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 27
;   LENGTH: 3010
;   TYPE: PRT
;   ORGANISM: Hepatitis C virus
US-09-539-601-27

  Query Match             74.5%;  Score 12001.5;  DB 4;  Length 3010;
  Best Local Similarity   71.6%;  Pred. No. 0;
  Matches 2177;  Conservative  354;  Mismatches  473;  Indels   35;  Gaps    7;

Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
Db          1 MGTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
Db         61 RRQPIPKARQPEGRAWAQPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPTDPRRRSRNLG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
Db        121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
Db        181 LLSCLTIPASAYEVRNVSGVYHVTNDCSNASIVYEAADMIMHTPGCVPCVRENNSSRCWV 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
Db        241 ALTPTLAARNASVPTTTIRRHVDLLVGAAALCSAMYVGDLCGSVFLVAQLFTFSPRRHET 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
Db        301 VQDCNCSIYPGHVTGHRMAWDMMMNWSPTAALVVSQLLRIPQAVVDMVAGAHWGVLAGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
Db        361 YYSMVGNWAKVLIVMLLFAGVDGGTYVTGGTMAKNTLGITSLFSPGSSQKIQLVNTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
Db        421 HINRTALNCNDSLNTGFLAALFYVHKFNSSGCPERMASCSPIDAFAQGWGPITY--NESH 478

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
Db        479 SSDQRPYCWHYAPRPCGIVPAAQVCGPVYCFTPSPVVVGTTDRFGVPTYSWGENETDVLL 538

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
Db        539 LNNTRPPQGNWFGCTWMNSTGFTKTCGGPPCNIGGIGNKT--LTCPTDCFRKHPEATYTK 596

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
Db        597 CGSGPWLTPRCLVHYPYRLWHYPCTVNFTIFKVRMYVGGVEHRLEAACNWTRGERCNLED 656

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
Db        657 RDRSELSPLLLSTTEWQVLPCSFTTLPALSTGLIHLHQNVVDVQYLYGIGSAVVSFAIKW 716

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
Db        717 EYVLLLFLLLADARVCACLWMMLLIAQAEAALENLVVLNAASVAGAHGILSFLVFFCAAW 776

Qy        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
Db        777 YIKGRLVPGAAYALYGVWPLLLLLLALPPRAYAMDREMAASCGGAVFVGLILLTLSPHYK 836

Qy        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
Db        837 LFLARLIWWLQYFITRAEAHLQVWIPPLNVRGGRDAVILLTCAIHPELIFTITKILLAIL 896

Qy        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
Db        897 GPLMVLQAGITKVPYFVRAHGLIRACMLVRKVAGGHYVQMALMKLAALTGTYVYDHLTPL 956

Qy        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
Db        957 RDWAHAGLRDLAVAVEPVVFSDMETKVITWGADTAACGDIILGLPVSARRGREIHLGPAD 1016

Qy       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080
Db       1017 SLEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDRNQVEGEVQVVSTATQSFLATCVNG 1076

Qy       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
Db       1077 VCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRH 1136

Qy       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200
Db       1137 ADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFV 1196

Qy       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260
Db       1197 PVESMETTMRSPVFTDNSSPPAVPQTFQVAHLHAPTGSGKSTKVPAAYAAQGYKVLVLNP 1256

Qy       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320
Db       1257 SVAATLGFGAYMSKAHGIDPNIRIGVRTITTGAPITYSTYGKFLADGGCSGGAYDIIICD 1316

Qy       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380
Db       1317 ECHSTDSTTILGIGTVLDQAETAGARLVVLATATPPGSVTVPHPNIEEVALSSTGEIPFY 1376

Qy       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440
Db       1377 GKAIPIETIKGGRHLIFCHSKKKCDELAAKLSGLGLNAVAYYRGLDVSVIPTSGDVIVVA 1436

Qy       1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500
Db       1437 TDALMTGFTGDFDSVIDCNTCVTQTVDFSLDPTFTIETTTVPQDAVSRSQRRGRTGRGRM 1496

Qy       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560
Db       1497 GIYRFVTPGERPSGMFDSSVLCECYDAGCAWYELTPAETSVRLRAYLNTPGLPVCQDHLE 1556

Qy       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
Db       1557 FWESVFTGLTHIDAHFLSQTKQAGDNFPYLVAYQATVCARAQAPPPSWDQMWKCLIRLKP 1616
```

```
Qy       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680
Db       1617 TLHGPTPLLYRLGAVQNEVTTTHPITKYIMACMSADLEVVTSTWVLVGGVLAALAAYCLT 1676

Qy       1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740
Db       1677 TGSVVIVGRIILSGKPAIIPDREVLYREFDEMEECASHLPYIEQGMQLAEQFKQKAIGLL 1736

Qy       1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
Db       1737 QTATKQAEAAAAPVVESKWRTIEAFWAKHMWNFISGIQYLAGLSTLPGNPAIASLMAFTAS 1796

Qy       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860
Db       1797 ITSPLTTQHTLLFNILGGWVAAQLAPPSAASAFVGAGIAGAAVGSIGLGKVLVDILAGYG 1856

Qy       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920
Db       1857 AGVAGALVAFKVMSGEMPSTEDLVNLLPAILSPGALVVGVVCAAILRRHVGPGEGAVQWM 1916

Qy       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980
Db       1917 NRLIAFASRGNHVSPTHYVPESDAAARVTQILSSLTITQLLKRLHQWINEDCSTPCSGSW 1976

Qy       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040
Db       1977 LRDVWDWICTVLTDFKTWLQSKLLPRLPGVPFFSCQRGYKGVWRGDGIMQTTCPCGAQIT 2036

Qy       2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100
Db       2037 GHVKNGSMRIVGPRTCSNTWHGTFPINAYTTGPCTPSPAPNYSRALWRVAAEEYVEVTRV 2096

Qy       2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160
Db       2097 GDFHYVTGMTTDDVKCPCQVPAPEFFTEVDGVRLHRYAPACKPLLREEVTFLVGLNQYLV 2156

Qy       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220
Db       2157 GSQLPCEPEPDVAVLTSMLTDPSHITAETAKRRLARGSPPPLASSSASQLSAPSLKATCT 2216

Qy       2221 THGKAYDVDMVDANLF----MGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYML 2276
Db       2217 TRHDSPDADLIEANLLWRQEMGGNITRVESENKVVILDSFEPLQAEEDEREVSVPAEILR 2276

Qy       2277 PKKRFPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLS 2336
Db       2277 RSRKFPRAMPIWARPDYNPPLLESWKDPDYVPPVVHGCPLPPAKAPPIPPSRRKRTVVLS 2336

Qy       2337 EDSIGDALQQLAIKSFGQPPPSG-DSGLSTGAGAADSGSQTPPDELALSETGSISSMPPL 2395
Db       2337 ESTVSSALAELATETFGSSESSAVDSGTATA-----SPDQPSDDGDAGSDVESYSSMPPL 2391

Qy       2396 EGELGDPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEE-DDSVVCCSMSYSWTGALIT 2454
Db       2392 EGEPGDPDL--------------------SDGSWSTVSEEASEDVVCCSMSYTWTGALIT 2431

Qy       2455 PCSPEEEKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIK 2514
Db       2432 PCAAEETKLPINALSNSLLRHHNLVYATTSRSASLRQKKVTFDRLQVLDDHYRDVLKEMK 2491

Qy       2515 LAASKVTARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSET 2574
Db       2492 AKASTVKAKLLSVEEACKLTPPHSARSKFGYGAKDVRNLSSKAVNHIRSVWKDLLEDTET 2551

Qy       2575 PIPTTIMAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYG 2634
Db       2552 PIDTTIMAKNEVFCVQPEKGGRKPARLIVFPDLGVRVCEKMALYDVVSTLPQAVMGSSYG 2611

Qy       2635 FQYSPAQRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTA 2694
Db       2612 FQYSPGQRVEFLVNAWKAKKCPMGFAYDTRCFDSTVTENDIRVEESIYQCCDLAPEARQA 2671

Qy       2695 IHSLTERLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPT 2754
Db       2672 IRSLTERLYIGGPLTNSKGQNCGYRRCRASGVLTTSCGNTLTCYLKAAAACRAAKLQDCT 2731

Qy       2755 MLVCGDDLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSV 2814
Db       2732 MLVCGDDLVVICESAGTQEDEASLRAFTEAMTRYSAPPGDPPKPEYDLELITSCSSNVSV 2791

Qy       2815 ALGPQGRRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSI 2874
Db       2792 AHDASGKRVYYLTRDPTTPLARAAWETARHTPVNSWLGNIIMYAPTLWARMILMTHFFSI 2851

Qy       2875 LMAQDTLDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKL 2934
Db       2852 LLAQEQLEKALDCQIYGACYSIEPLDLPQIIQRLHGLSAFSLHSYSPGEINRVASCLRKL 2911

Qy       2935 GAPPLRAWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFT 2994
Db       2912 GVPPLRVWRHRARSVRARLLSQGGRAATCGKYLFNWAVRTKLKLTPIPAASQLDLSSWFV 2971

Qy       2995 VGAGGGDIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
Db       2972 AGYSGGDIYHSLSRARPRWFMWCLLLLSVGVGIYLLPNR 3010
```

RESULT 7
US-08-324-977-2
; Sequence 2, Application US/08324977
; Patent No. 5747339
; GENERAL INFORMATION:
; APPLICANT: OKAYAMA, Hiroto
; APPLICANT: FUKE, Isao
; APPLICANT: MORI, Chisato
; APPLICANT: TAKAMIZAWA, Akahisa
; APPLICANT: YOSHIDA, Iwao
; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
; TITLE OF INVENTION: cDNA AND ANTIGEN POLYPEPTIDE
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Armstrong, Westerman, Hattori, McLeland &
; ADDRESSEE: Naughton
; STREET: 1725 K St. N.W. Suite 1000
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 in, 1.44Mb
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/324,977
; FILING DATE: 18-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-167466
; FILING DATE: 25-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-230921
; FILING DATE: 31-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-305605
; FILING DATE: 09-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/099,706
; FILING DATE: 30-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/769,996
; FILING DATE: 02-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/635,451
; FILING DATE: 28-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Stevens-Smith, Theresa M.

```
;      REGISTRATION NUMBER:  36,281
;      REFERENCE/DOCKET NUMBER:  900703D
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 659-2930
;      TELEFAX:  (202) 887-0357
;      TELEX:  440142
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3010 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-324-977-2

  Query Match             74.3%;  Score 11984.5;  DB 1;  Length 3010;
  Best Local Similarity   71.5%;  Pred. No. 0;
  Matches 2175;  Conservative  352;  Mismatches  472;  Indels   45;  Gaps    8;

Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRAPRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              |||||||  ||   |::| :|||||||||||||||||||||||||||||||| ||| ||||:|
Db         61 RRQPIPKARRPEGRTWAQPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPTDPRRRSRNLG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
              |||||||||||||||||:|||||| ||||||||||||||||:||||||||||||||||||
Db        121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
              ||||:||| || || |:|  | |||||:| || ::    ::| |||||| : ||:|:||:
Db        181 LLSCLTTPASAYEVHNVSGIYHVTNDCSNASIVYEAADLIMHTPGCVPCVREGNSSRCWV 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
               ::| :| :     |  :| |:|::| :|  |||:|||||| | | :|:|  ||: |
Db        241 ALTPTLAARNVTIPTTTIRRHVDLLVGAAAFCSAMYVGDLCGSVFLVSQLFTFSPRRHVT 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
              :|||||||||| ::|||||||||||||||  ::::  :|:|: ::|:::|||||: |||
Db        301 LQDCNCSIYPGHVSGHRMAWDMMMNWSPTTALVVSQLLRIPQAVVDMVAGAHWGVLAGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
              |:|| | ||||:::::|| |||| || ||: |:|| || |:|  || |||||:||||||
Db        361 YYSMAGNWAKVLIVMLLFAGVDGDTHVTGGAQAKTTNRLVSMFASGPSQKIQLINTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
              |||||||||||| |||:|:|||||||||||||||: ||:|: |  ||| : | :  ::
Db        421 HINRTALNCNDSLQTGFLAALFYTHSFNSSGCPERMAQCRTIDKFDQGWGPITYAE--SS 478

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
              | ||||||||| || :|  | |||||||||||||||||||| | ||| |||||||| |
Db        479 RSDQRPYCWHYPPPQCTIVPASEVCGPVYCFTPSPVVVGTTDRFGVPTYRWGENETDVLL 538

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
              ||:|||| |:||||||||:|:|||||| |||     |  | ||||||||||: || |
Db        539 LNNTRPPQGNWFGCTWMNSTGFTKTCGGPPCNIGGVGN--NTLTCPTDCFRKHPEATYTK 596

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
              ||||||||||::|||||||||||||:||||:|||||||||| ||||:|||:||:|||
Db        597 CGSGPWLTPRCMVDYPYRLWHYPCTVNFTIFKVRMYVGGVEHRLNAACNWTRGERCDLED 656

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
              ||| :||||| ||||| :||||:: |||||||||:||||||||||::||:  |:  : ::|
Db        657 RDRPELSPLLLSTTEWQVLPCSFTTLPALSTGLIHLHQNIVDVQYLYGIGSAVVSFAIKW 716

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
              |:|:||||||||||||||||::|: ||||||| ||:|::|| |  :| | |::|| |||
Db        717 EYVLLLFLLLADARVCACLWMMLLIAQAEAALENLVVLNSASVAGAHGILSFLVFFCAAW 776

Qy        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
              |||||:|| |||:| |:|   |||||||  :||| |  :    | |: | : | ||:| ||
Db        777 YIKGRLVPGATYALYGVWPLLLLLLALPPRAYAMDREMAASCGGAVFVGLVLLTLSPYYK 836

Qy        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
              |:| :||| |  |  || :  | ||:  ||||| ||  :   :| ::|||||  |:|:|
Db        837 VFLARLIWWLQYFTTRAEADLHVWIPPLNARGGRDAIILLMCAVHPELIFDITKLLIAIL 896

Qy        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
              ||  :|:  :|||||||||  |:  | : | :||| ||||| : ||  ||||||:||||:
Db        897 GPLMVLQAGITRVPYFVRAQGLIHACMLVRKVAGGHYVQMAFMKLGALTGTYIYNHLTPL 956

Qy        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
              ||  :|||||||||||::|| || |:| |||:|||||||: ||||||| |:|:||||||
Db        957 RDWPRAGLRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGKEILLGPAD 1016

Qy       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080
                 :|  ||||||||:|||||||| |: |:|||||  :  ||:||:|| ||||| | ::|
Db       1017 SLEGRGLRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNG 1076

Qy       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
              | ||||||||:|||  :||:||||:: : |||||| ||| :|| |||||: |||||||:
Db       1077 VCWTVYHGAGSKTLAAPKGPITQMYTNVDQDLVGWPKPPGARSLTPCTCGSSDLYLVTRH 1136

Qy       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200
              ||||| ||||| ||:||||||:| ||||||||:||| |||||:||||||:|||||::||:
Db       1137 ADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPFGHAVGIFRAAVCTRGVAKAVDFV 1196

Qy       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260
              |||:::   ||| |:|||:||||||::|| :|||||||||||| ||||||||||||||
Db       1197 PVESMETTMRSPVFTDNSSPPAVPQSFQVAHLHAPTGSGKSTKVPAAYAAQGYKVLVLNP 1256

Qy       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320
              |||||||||||:||||||:||||:|||||:|||||||||||:||||||||||||||||||
Db       1257 SVAATLGFGAYMSKAHGIDPNIRTGVRTITTGAPVTYSTYGKFLADGGCSGGAYDIIICD 1316

Qy       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380
              |||: |||||||||||||||||| || |||||||||| |||||||||   ||||||
Db       1317 ECHSTDSTTILGIGTVLDQAETAGARLVVLATATPPGSVTVPHPNIEEVALSNTGEIPFY 1376

Qy       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440
              |:|||:  |:|||||||||||||||||| | |:|:|:|||||||||||||| |||||||
Db       1377 GKAIPIEAIRGGRHLIFCHSKKKCDELAAKLSGLGINAVAYYRGLDVSVIPTIGDVVVVA 1436

Qy       1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500
              |||||||||||||||||| ||| |||||||||| |  ||||||||||||||||||||
Db       1437 TDALMTGYTGDFDSVIDCNTCVTQTVDFSLDPTFTIETTTVPQDAVSRSQRRGRTGRGRR 1496

Qy       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560
              ||||:|: ||| |||||| |||||||| ||||||||:|:|||||| ||||||||||||
Db       1497 GIYRFVTPGERPSGMFDSSVLCECYDAGCAWYELTPAETSVRLRAYLNTPGLPVCQDHLE 1556

Qy       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
              |||:|||||||||||||||:|:|| || |||||||||:|||||| ||||| ||||
Db       1557 FWESVFTGLTHIDAHFLSQTKQAGDNFPYLVAYQATVCARAQAPPPSWDQMWKCLIRLKP 1616

Qy       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680
              || |||||||||:| |||||||:|||  || |||||:|||||| ||||||:|||||
Db       1617 TLHGPTPLLYRLGAVQNEVTLTHPITKYIMACMSADLEVVTSTWVLVGGVLAALAAYCLT 1676

Qy       1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740
              || | |:||: :: |  : ||:|:||: |||||||||    ||:| ::||  | |  |||
Db       1677 TGSVVIVGRIILSGRPAIVPDRELLYQEFDEMEECASHLPYIEQGMQLAEQFKQKALGLL 1736

Qy       1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
              | |:|||:   | |:: |  :| |||||||||||||||||||||||||:||:|||:|:
Db       1737 QTATKQAEAAAPVVESKWRALETFWAKHMWNFISGIQYLAGLSTLPGNPAIASLMAFTAS 1796

Qy       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860
              :||||:| :|:| |||||:|:|:|||: |: || :|: ||||||||||||||||||||
Db       1797 ITSPLTTQSTLLFNILGGWVAAQLAPPSAASAFVGAGIAGAAVGSIGLGKVLVDILAGYG 1856

Qy       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920
              ||::|||||||:|||| || ||:||||| ||||||||||:|||||||||||||||||||
Db       1857 AGVAGALVAFKVMSGEMPSTEDLVNLLPAILSPGALVVGVVCAAILRRHVGPGEGAVQWM 1916
```

```
Qy       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980
Db       1917 NRLIAFASRGNHVSPTHYVPESDAAARVTQILSSLTITQLLKRLHQWINEDCSTPCSGSW 1976

Qy       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040
Db       1977 LRDVWDWICTVLTDFKTWLQSKLLPQLPGVPFFSCQRGYKGVWRGDGIMQTTCPCGAQIT 2036

Qy       2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100
Db       2037 GHVKNGSMRIVGPKTCSNTWHGTFPINAYTTGPCTPSPAPNYSRALWRVAAEEYVEVTRV 2096

Qy       2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160
Db       2097 GDFHYVTGMTTDNVKCPCQVPAPEFFSEVDGVRLHRYAPACRPLLREEVTFQVGLNQYLV 2156

Qy       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220
Db       2157 GSQLPCEPEPDVAVLTSMLTDPSHITAETAKRRLARGSPPSLASSSASQLSAPSLKATCT 2216

Qy       2221 THGKAYDVDMVDANLF----MGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYML 2276
Db       2217 THHVSPDADLIEANLLWRQEMGGNITRVESENKVVVLDSFDPLRAEEDEREVSVPAEILR 2276

Qy       2277 PKKRFPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLS 2336
Db       2277 KSKKFPAAMPIWARPDYNPPLLESWKDPDYVPPVVHGCPLPPIKAPPIPPPRRKRTVVLT 2336

Qy       2337 EDSIGDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTP-PDELA-----LSETGSIS 2390
Db       2337 ESSVSSALAELATKTFG----------SSESSAVDSGTATALPDQASDDGDKGSDVESYS 2386

Qy       2391 SMPPLEGELGDPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEE-DDSVVCCSMSYSWT 2449
Db       2387 SMPPLEGEPGDPDL--------------------SDGSWSTVSEEASEDVVCCSMSYTWT 2426

Qy       2450 GALITPCSPEEEKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSV 2509
Db       2427 GALITPCAAEESKLPINALSNSLLRHHNMVYATTSRSAGLRQKKVTFDRLQVLDDHYRDV 2486

Qy       2510 LKDIKLAASKVTARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLL 2569
Db       2487 LKEMKAKASTVKAKLLSVEEACKLTPPHSAKSKFGYGAKDVRNLSSKAVNHIHSVWKDLL 2546

Qy       2570 EDSETPIPTTIMAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVM 2629
Db       2547 EDTVTPIDTTIMAKNEVFCVQPEKGGRKPARLIVFPDLGVRVCEKMALYDVVSTLPQVVM 2606

Qy       2630 GASYGFQYSPAQRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPE 2689
Db       2607 GSSYGFQYSPGQRVEFLVNTWKSKKNPMGFSYDTRCFDSTVTENDIRVEESIYQCCDLAP 2666

Qy       2690 EAHTAIHSLTERLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAG 2749
Db       2667 EARQAIKSLTERLYIGGPLTNSKGQNCGYRRCRASGVLTTSCGNTLTCYLKASAACRAAK 2726

Qy       2750 IIAPTMLVCGDDLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCS 2809
Db       2727 LQDCTMLVNGDDLVVICESAGTQEDAASLRVFTEAMTRYSAPPGDPPQPEYDLELITSCS 2786

Qy       2810 SNVSVALGPQGRRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMT 2869
Db       2787 SNVSVAHDASGKRVYYLTRDPTTPLARAAWETARHTPVNSWLGNIIMYAPTLWARMILMT 2846

Qy       2870 HFFSILMAQDTLDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVAS 2929
Db       2847 HFFSILLAQEQLEKALDCQIYGACYSIEPLDLPQIIERLHGLSAFSLHSYSPGEINRVAS 2906

Qy       2930 ALRKLGAPPLRAWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDL 2989
Db       2907 CLRKLGVPPLRVWRHRARSVRARLLSQGGRAATCGKYLFNWAVKTKLKLTPIPAASRLDL 2966

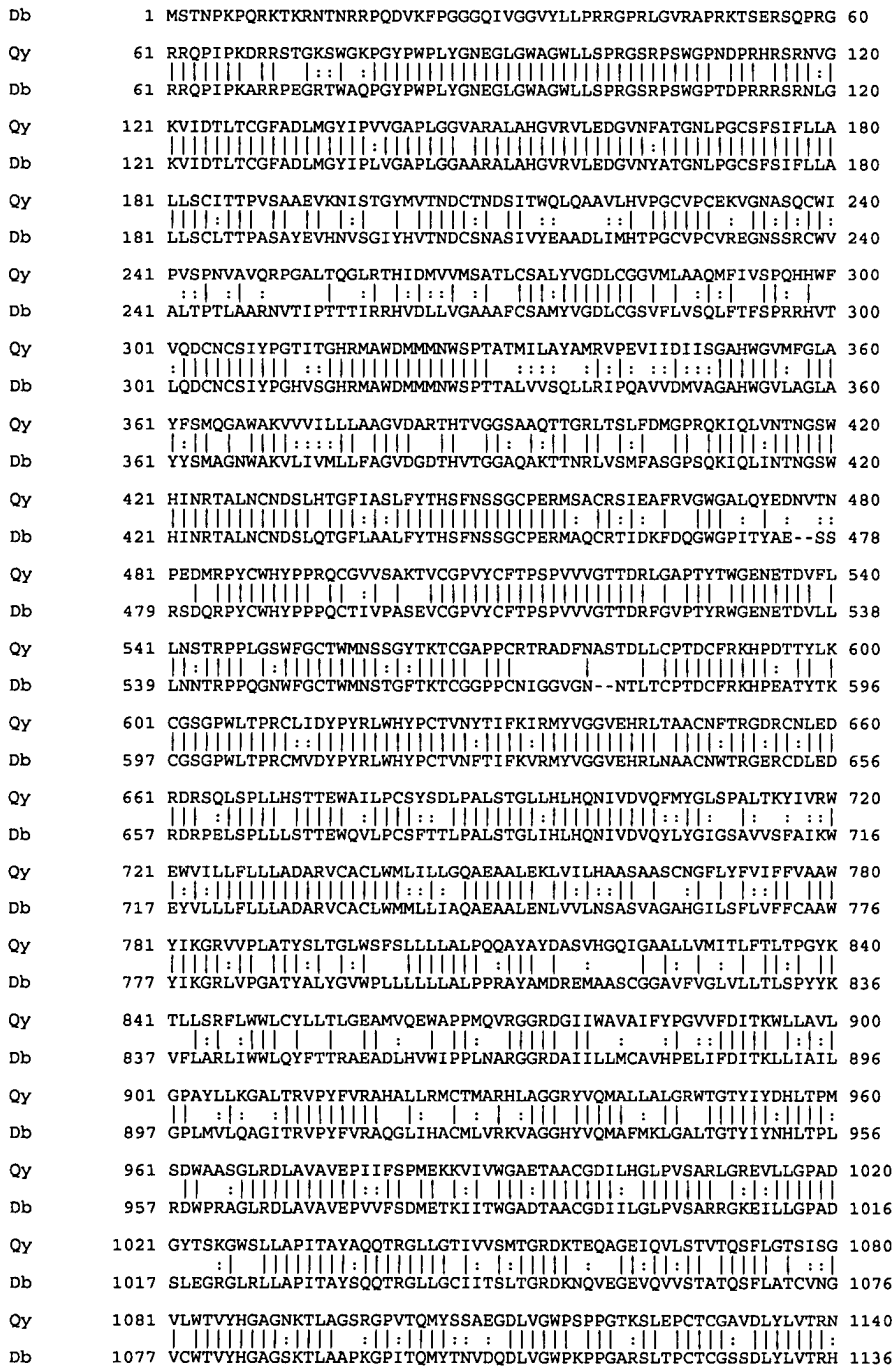
Qy       2990 SSWFTVGAGGGDIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
```

```
Db       2967 SGWFVAGYSGGDIYHSLSRARPRWFMLCLLLLSVGVGIYLLPNR 3010
```

```
RESULT 8
US-08-324-977-14
; Sequence 14, Application US/08324977
; Patent No. 5747339
;  GENERAL INFORMATION:
;    APPLICANT:  OKAYAMA, Hiroto
;    APPLICANT:  FUKE, Isao
;    APPLICANT:  MORI, Chisato
;    APPLICANT:  TAKAMIZAWA, Akahisa
;    APPLICANT:  YOSHIDA, Iwao
;    TITLE OF INVENTION:  NON-A, NON-B HEPATITIS VIRUS GENOMIC
;    TITLE OF INVENTION:  cDNA AND ANTIGEN POLYPEPTIDE
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Armstrong, Westerman, Hattori, McLeland &
;      ADDRESSEE:  Naughton
;      STREET:  1725 K St. N.W. Suite 1000
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20006
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 in, 1.44Mb
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS, Version 5.0
;      SOFTWARE:  ASCII
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/324,977
;      FILING DATE:  18-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 2-167466
;      FILING DATE:  25-JUN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 2-230921
;      FILING DATE:  31-AUG-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 2-305605
;      FILING DATE:  09-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/099,706
;      FILING DATE:  30-JUL-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/769,996
;      FILING DATE:  02-OCT-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/635,451
;      FILING DATE:  28-DEC-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stevens-Smith, Theresa M.
;      REGISTRATION NUMBER:  36,281
;      REFERENCE/DOCKET NUMBER:  900703D
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 659-2930
;      TELEFAX:  (202) 887-0357
;      TELEX:  440142
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3010 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-324-977-14

  Query Match             74.3%;  Score 11984.5;  DB 1;  Length 3010;
  Best Local Similarity   71.5%;  Pred. No. 0;
  Matches 2175;  Conservative  352;  Mismatches  472;  Indels   45;  Gaps    8;

Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
```


Db 1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRAPRKTSERSQPRG 60
Qy 61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
Db 61 RRQPIPKARRPEGRTWAQPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPTDPRRRSRNLG 120
Qy 121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
Db 121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180
Qy 181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
Db 181 LLSCLTTPASAYEVHNVSGIYHVTNDCSNASIVYEAADLIMHTPGCVPCVREGNSSRCWV 240
Qy 241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
Db 241 ALTPTLAARNVTIPTTTIRRHVDLLVGAAAFCSAMYVGDLCGSVFLVSQLFTFSPRRHVT 300
Qy 301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
Db 301 LQDCNCSIYPGHVSGHRMAWDMMMNWSPTTALVVSQLLRIPQAVVDMVAGAHWGVLAGLA 360
Qy 361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
Db 361 YYSMAGNWAKVLIVMLLFAGVDGDTHVTGGAQAKTTNRLVSMFASGPSQKIQLINTNGSW 420
Qy 421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
Db 421 HINRTALNCNDSLQTGFLAALFYTHSFNSSGCPERMAQCRTIDKFDQGWGPITYAE--SS 478
Qy 481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
Db 479 RSDQRPYCWHYPPPQCTIVPASEVCGPVYCFTPSPVVVGTTDRFGVPTYRWGENETDVLL 538
Qy 541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
Db 539 LNNTRPPQGNWFGCTWMNSTGFTKTCGGPPCNIGGVGN--NTLTCPTDCFRKHPEATYTK 596
Qy 601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
Db 597 CGSGPWLTPRCMVDYPYRLWHYPCTVNFTIFKVRMYVGGVEHRLNAACNWTRGERCDLED 656
Qy 661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
Db 657 RDRPELSPLLLSTTEWQVLPCSFTTLPALSTGLIHLHQNIVDVQYLYGIGSAVVSFAIKW 716
Qy 721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
Db 717 EYVLLLFLLLADARVCACLWMMLLIAQAEAALENLVVLNSASVAGAHGILSFLVFFCAAW 776
Qy 781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
Db 777 YIKGRLVPGATYALYGVWPLLLLLLALPPRAYAMDREMAASCGGAVFVGLVLLTLSPYYK 836
Qy 841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
Db 837 VFLARLIWWLQYFTTRAEADLHVWIPPLNARGGRDAIILLMCAVHPELIFDITKLLIAIL 896
Qy 901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
Db 897 GPLMVLQAGITRVPYFVRAQGLIHACMLVRKVAGGHYVQMAFMKLGALTGTYIYNHLTPL 956
Qy 961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
Db 957 RDWPRAGLRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGKEILLGPAD 1016
Qy 1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080
Db 1017 SLEGRGLRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNG 1076
Qy 1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
Db 1077 VCWTVYHGAGSKTLAAPKGPITQMYTNVDQDLVGWPKPPGARSLTPCTCGSSDLYLVTRH 1136
Qy 1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200
Db 1137 ADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPFGHAVGIFRAAVCTRGVAKAVDFV 1196
Qy 1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260
Db 1197 PVESMETTMRSPVFTDNSSPPAVPQSFQVAHLHAPTGSGKSTKVPAAYAAQGYKVLVLNP 1256
Qy 1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320
Db 1257 SVAATLGFGAYMSKAHGIDPNIRTGVRTITTGAPVTYSTYGKFLADGGCSGGAYDIIICD 1316
Qy 1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380
Db 1317 ECHSTDSTTILGIGTVLDQAETAGARLVVLATATPPGSVTVPHPNIEEVALSNTGEIPFY 1376
Qy 1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440
Db 1377 GKAIPIEAIRGGRHLIFCHSKKKCDELAAKLSGLGINAVAYYRGLDVSVIPTIGDVVVVA 1436
Qy 1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500
Db 1437 TDALMTGYTGDFDSVIDCNTCVTQTVDFSLDPTFTIETTTVPQDAVSRSQRRGRTGRGRR 1496
Qy 1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560
Db 1497 GIYRFVTPGERPSGMFDSSVLCECYDAGCAWYELTPAETSVRLRAYLNTPGLPVCQDHLE 1556
Qy 1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
Db 1557 FWESVFTGLTHIDAHFLSQTKQAGDNFPYLVAYQATVCARAQAPPPSWDQMWKCLIRLKP 1616
Qy 1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680
Db 1617 TLHGPTPLLYRLGAVQNEVTLTHPITKYIMACMSADLEVVTSTWVLVGGVLAALAAYCLT 1676
Qy 1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740
Db 1677 TGSVVIVGRIILSGRPAIVPDRELLYQEFDEMEECASHLPYIEQGMQLAEQFKQKALGLL 1736
Qy 1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
Db 1737 QTATKQAEAAAPVVESKWRALETFWAKHMWNFISGIQYLAGLSTLPGNPAIASLMAFTAS 1796
Qy 1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860
Db 1797 ITSPLTTQSTLLFNILGGWVAAQLAPPSAASAFVGAGIAGAAVGSIGLGKVLVDILAGYG 1856
Qy 1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920
Db 1857 AGVAGALVAFKVMSGEMPSTEDLVNLLPAILSPGALVVGVVCAAILRRHVGPGEGAVQWM 1916
Qy 1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980
Db 1917 NRLIAFASRGNHVSPTHYVPESDAAARVTQILSSLTITQLLKRLHQWINEDCSTPCSGSW 1976
Qy 1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040
Db 1977 LRDVWDWICTVLTDFKTWLQSKLLPQLPGVPFFSCQRGYKGVWRGDGIMQTTCPCGAQIT 2036
Qy 2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100
Db 2037 GHVKNGSMRIVGPKTCSNTWHGTFPINAYTTGPCTPSPAPNYSRALWRVAAEEYVEVTRV 2096
Qy 2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160
Db 2097 GDFHYVTGMTTDNVKCPCQVPAPEFFSEVDGVRLHRYAPACRPLLREEVTFQVGLNQYLV 2156
Qy 2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220
Db 2157 GSQLPCEPEPDVAVLTSMLTDPSHITAETAKRRLARGSPPSLASSSASQLSAPSLKATCT 2216

```
Qy       2221 THGKAYDVDMVDANLF----MGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYML 2276
Db       2217 THHVSPDADLIEANLLWRQEMGGNITRVESENKVVVLDSFDPLRAEEDEREVSVPAEILR 2276

Qy       2277 PKKRFPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLS 2336
Db       2277 KSKKFPAAMPIWARPDYNPPLLESWKDPDYVPPVVHGCPLPPIKAPPIPPPRRKRTVVLT 2336

Qy       2337 EDSIGDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTP-PDELA-----LSETGSIS 2390
Db       2337 ESSVSSALAELATKTFG----------SSESSAVDSGTATALPDQASDDGDKGSDVESYS 2386

Qy       2391 SMPPLEGELGDPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEE-DDSVVCCSMSYSWT 2449
Db       2387 SMPPLEGEPGDPDL-------------------SDGSWSTVSEEASEDVVCCSMSYTWT 2426

Qy       2450 GALITPCSPEEEKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSV 2509
Db       2427 GALITPCAAEESKLPINALSNSLLRHHNMVYATTSRSAGLRQKKVTFDRLQVLDDHYRDV 2486

Qy       2510 LKDIKLAASKVTARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLL 2569
Db       2487 LKEMKAKASTVKAKLLSVEEACKLTPPHSAKSKFGYGAKDVRNLSSKAVNHIHSVWKDLL 2546

Qy       2570 EDSETPIPTTIMAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVM 2629
Db       2547 EDTVTPIDTTIMAKNEVFCVQPEKGGRKPARLIVFPDLGVRVCEKMALYDVVSTLPQVVM 2606

Qy       2630 GASYGFQYSPAQRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPE 2689
Db       2607 GSSYGFQYSPGQRVEFLVNTWKSKKNPMGFSYDTRCFDSTVTENDIRVEESIYQCCDLAP 2666

Qy       2690 EAHTAIHSLTERLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAG 2749
Db       2667 EARQAIKSLTERLYIGGPLTNSKGQNCGYRRCRASGVLTTSCGNTLTCYLKASAACRAAK 2726

Qy       2750 IIAPTMLVCGDDLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCS 2809
Db       2727 LQDCTMLVNGDDLVVICESAGTQEDAASLRVFTEAMTRYSAPPGDPPQPEYDLELITSCS 2786

Qy       2810 SNVSVALGPQGRRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMT 2869
Db       2787 SNVSVAHDASGKRVYYLTRDPTTPLARAAWETARHTPVNSWLGNIIMYAPTLWARMILMT 2846

Qy       2870 HFFSILMAQDTLDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVAS 2929
Db       2847 HFFSILLAQEQLEKALDCQIYGACYSIEPLDLPQIIERLHGLSAFSLHSYSPGEINRVAS 2906

Qy       2930 ALRKLGAPPLRAWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDL 2989
Db       2907 CLRKLGVPPLRVWRHRARSVRARLLSQGGRAATCGKYLFNWAVKTKLKLTPIPAASRLDL 2966

Qy       2990 SSWFTVGAGGGDIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
Db       2967 SGWFVAGYSGGDIYHSLSRARPRWFMLCLLLLSVGVGIYLLPNR 3010


RESULT 9
US-08-384-616-2
; Sequence 2, Application US/08384616
; Patent No. 5847101
;  GENERAL INFORMATION:
;    APPLICANT:  OKAYAMA, Hiroto
;    APPLICANT:  FUKE, Isao
;    APPLICANT:  MORI, Chisato
;    APPLICANT:  TAKAMIZAWA, Akahisa
;    APPLICANT:  YOSHIDA, Iwao
;    TITLE OF INVENTION:  NON-A, NON-B HEPATITIS VIRUS GENOMIC
;    TITLE OF INVENTION:  cDNA AND ANTIGEN POLYPEPTIDE
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Armstrong, Westerman, Hattori, McLeland &
;      ADDRESSEE:  Naughton
;      STREET:  1725 K St. N.W. Suite 1000
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20006
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 in, 1.44Mb
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS, Version 5.0
;      SOFTWARE:  ASCII
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/384,616
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/769,996
;      FILING DATE:  02-OCT-1991
;      APPLICATION NUMBER:  JP 2-167466
;      FILING DATE:  25-JUN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 2-230921
;      FILING DATE:  31-AUG-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 2-305605
;      FILING DATE:  09-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/635,451
;      FILING DATE:  28-DEC-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stevens-Smith, Theresa M.
;      REGISTRATION NUMBER:  36,281
;      REFERENCE/DOCKET NUMBER:  900703B
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 659-2930
;      TELEFAX:  (202) 887-0357
;      TELEX:  440142
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3010 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-384-616-2

  Query Match             74.3%;  Score 11984.5;  DB 2;  Length 3010;
  Best Local Similarity   71.5%;  Pred. No. 0;
  Matches 2175;  Conservative  352;  Mismatches  472;  Indels   45;  Gaps    8;

Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRAPRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
Db         61 RRQPIPKARRPEGRTWAQPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPTDPRRRSRNLG 120

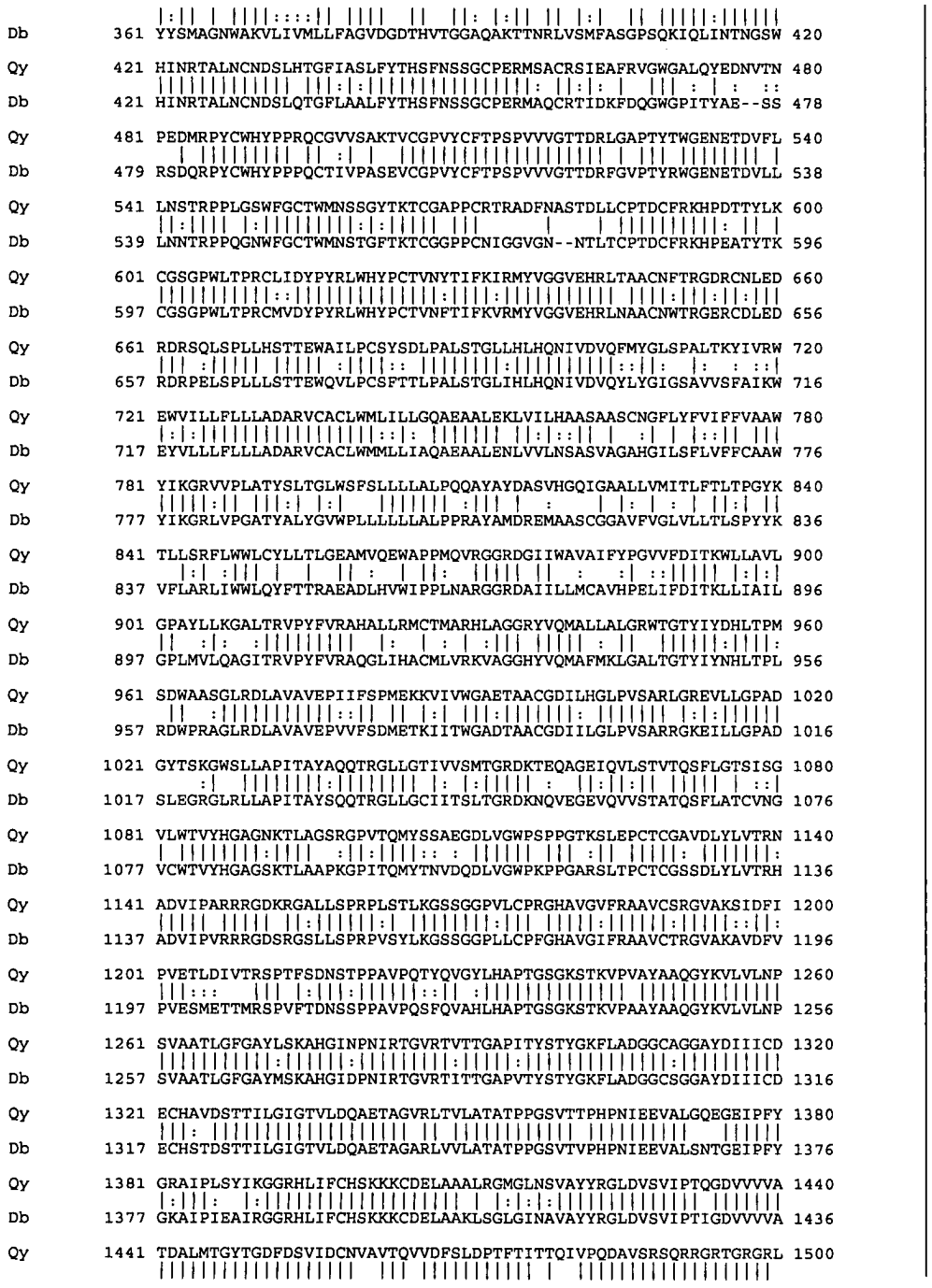
Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
Db        121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
Db        181 LLSCLTTPASAYEVHNVSGIYHVTNDCSNASIVYEAADLIMHTPGCVPCVREGNSSRCWV 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
Db        241 ALTPTLAARNVTIPTTTIRRHVDLLVGAAAFCSAMYVGDLCGSVFLVSQLFTFSPRRHVT 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
Db        301 LQDCNCSIYPGHVSGHRMAWDMMMNWSPTTALVVSQLLRIPQAVVDMVAGAHWGVLAGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
```



```
Qy       2570 EDSETPIPTTIMAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVM 2629
Db       2547 EDTVTPIDTTIMAKNEVFCVQPEKGGRKPARLIVFPDLGVRVCEKMALYDVVSTLPQVVM 2606

Qy       2630 GASYGFQYSPAQRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPE 2689
Db       2607 GSSYGFQYSPGQRVEFLVNTWKSKKNPMGFSYDTRCFDSTVTENDIRVEESIYQCCDLAP 2666

Qy       2690 EAHTAIHSLTERLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAG 2749
Db       2667 EARQAIKSLTERLYIGGPLTNSKGQNCGYRRCRASGVLTTSCGNTLTCYLKASAACRAAK 2726

Qy       2750 IIAPTMLVCGDDLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCS 2809
Db       2727 LQDCTMLVNGDDLVVICESAGTQEDAASLRVFTEAMTRYSAPPGDPPQPEYDLELITSCS 2786

Qy       2810 SNVSVALGPQGRRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMT 2869
Db       2787 SNVSVAHDASGKRVYYLTRDPTTPLARAAWETARHTPVNSWLGNIIMYAPTLWARMILMT 2846

Qy       2870 HFFSILMAQDTLDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVAS 2929
Db       2847 HFFSILLAQEQLEKALDCQIYGACYSIEPLDLPQIIERLHGLSAFSLHSYSPGEINRVAS 2906

Qy       2930 ALRKLGAPPLRAWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDL 2989
Db       2907 CLRKLGVPPLRVWRHRARSVRARLLSQGGRAATCGKYLFNWAVKTKLKLTPIPAASRLDL 2966

Qy       2990 SSWFTVGAGGGDIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
Db       2967 SGWFVAGYSGGDIYHSLSRARPRWFMLCLLLLSVGVGIYLLPNR 3010
```

```
RESULT 10
US-08-384-616-14
; Sequence 14, Application US/08384616
; Patent No. 5847101
;  GENERAL INFORMATION:
;    APPLICANT:  OKAYAMA, Hiroto
;    APPLICANT:  FUKE, Isao
;    APPLICANT:  MORI, Chisato
;    APPLICANT:  TAKAMIZAWA, Akahisa
;    APPLICANT:  YOSHIDA, Iwao
;    TITLE OF INVENTION:  NON-A, NON-B HEPATITIS VIRUS GENOMIC
;    TITLE OF INVENTION:  cDNA AND ANTIGEN POLYPEPTIDE
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Armstrong, Westerman, Hattori, McLeland &
;      ADDRESSEE:  Naughton
;      STREET:  1725 K St. N.W. Suite 1000
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20006
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 in, 1.44Mb
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS, Version 5.0
;      SOFTWARE:  ASCII
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/384,616
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/769,996
;      FILING DATE:  02-OCT-1991
;      APPLICATION NUMBER:  JP 2-167466
;      FILING DATE:  25-JUN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 2-230921
;      FILING DATE:  31-AUG-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 2-305605
;      FILING DATE:  09-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/635,451
;      FILING DATE:  28-DEC-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stevens-Smith, Theresa M.
;      REGISTRATION NUMBER:  36,281
;      REFERENCE/DOCKET NUMBER:  900703B
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 659-2930
;      TELEFAX:  (202) 887-0357
;      TELEX:  440142
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3010 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-384-616-14

  Query Match             74.3%;  Score 11984.5;  DB 2;  Length 3010;
  Best Local Similarity   71.5%;  Pred. No. 0;
  Matches 2175;  Conservative  352;  Mismatches  472;  Indels   45;  Gaps    8;

Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRAPRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
Db         61 RRQPIPKARRPEGRTWAQPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPTDPRRRSRNLG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
Db        121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
Db        181 LLSCLTTPASAYEVHNVSGIYHVTNDCSNASIVYEAADLIMHTPGCVPCVREGNSSRCWV 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
Db        241 ALTPTLAARNVTIPTTTIRRHVDLLVGAAAFCSAMYVGDLCGSVFLVSQLFTFSPRRHVT 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
Db        301 LQDCNCSIYPGHVSGHRMAWDMMMNWSPTTALVVSQLLRIPQAVVDMVAGAHWGVLAGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
Db        361 YYSMAGNWAKVLIVMLLFAGVDGDTHVTGGAQAKTTNRLVSMFASGPSQKIQLINTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
Db        421 HINRTALNCNDSLQTGFLAALFYTHSFNSSGCPERMAQCRTIDKFDQGWGPITYAE--SS 478

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
Db        479 RSDQRPYCWHYPPPQCTIVPASEVCGPVYCFTPSPVVVGTTDRFGVPTYRWGENETDVLL 538

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
Db        539 LNNTRPPQGNWFGCTWMNSTGFTKTCGGPPCNIGGVGN--NTLTCPTDCFRKHPEATYTK 596

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
Db        597 CGSGPWLTPRCMVDYPYRLWHYPCTVNFTIFKVRMYVGGVEHRLNAACNWTRGERCDLED 656

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
Db        657 RDRPELSPLLLSTTEWQVLPCSFTTLPALSTGLIHLHQNIVDVQYLYGIGSAVVSFAIKW 716
```



```
Db       2847 HFFSILLAQEQLEKALDCQIYGACYSIEPLDLPQIIERLHGLSAFSLHSYSPGEINRVAS 2906

Qy       2930 ALRKLGAPPLRAWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDL 2989
              ||||| |||| |: |||:||| |:|:||||| ||:|||||||||||||||:| |  |||
Db       2907 CLRKLGVPPLRVWRHRARSVRARLLSQGGRAATCGKYLFNWAVKTKLKLTPIPAASRLDL 2966

Qy       2990 SSWFTVGAGGGDIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
              | ||  |  |||||||:||||||  :  |||| ||||::||| |
Db       2967 SGWFVAGYSGGDIYHSLSRARPRWFMLCLLLLSVGVGIYLLPNR 3010


RESULT 11
US-08-904-686A-2
; Sequence 2, Application US/08904686A
; Patent No. 5998130
;  GENERAL INFORMATION:
;    APPLICANT:  OKAYAMA, Hiroto
;    APPLICANT:  FUKE, Isao
;    APPLICANT:  MORI, Chisato
;    APPLICANT:  TAKAMIZAWA, Akahisa
;    APPLICANT:  YOSHIDA, Iwao
;    TITLE OF INVENTION:  NON-A, NON-B HEPATITIS VIRUS GENOMIC
;    TITLE OF INVENTION:  cDNA AND ANTIGEN POLYPEPTIDE
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Armstrong, Westerman, Hattori, McLeland &
;      ADDRESSEE:  Naughton
;      STREET:  1725 K St. N.W. Suite 1000
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20006
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 in, 1.44Mb
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS, Version 5.0
;      SOFTWARE:  ASCII
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/904,686A
;      FILING DATE:  01-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/324,977
;      FILING DATE:  18-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 2-167466
;      FILING DATE:  25-JUN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 2-230921
;      FILING DATE:  31-AUG-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 2-305605
;      FILING DATE:  09-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/099,706
;      FILING DATE:  30-JUL-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/769,996
;      FILING DATE:  02-OCT-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/635,451
;      FILING DATE:  28-DEC-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McLeland, Le-Nhung
;      REGISTRATION NUMBER:  31,541
;      REFERENCE/DOCKET NUMBER:  900703G
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 659-2930
;      TELEFAX:  (202) 887-0357
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3010 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear

;    MOLECULE TYPE:  protein
US-08-904-686A-2

  Query Match             74.3%;  Score 11984.5;  DB 2;  Length 3010;
  Best Local Similarity   71.5%;  Pred. No. 0;
  Matches 2175;  Conservative  352;  Mismatches  472;  Indels   45;  Gaps    8;

Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRAPRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              ||||||| ||  |::| :|||||||||||||||||||||||||||||| ||| ||||:|
Db         61 RRQPIPKARRPEGRTWAQPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPTDPRRRSRNLG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
              ||||||||||||||||||:||||||| |||||||||||||||||:|||||||||||||||
Db        121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
              ||||:||| || || |:|  | |||||:| || ::    ::| |||||| : ||:|:||:
Db        181 LLSCLTTPASAYEVHNVSGIYHVTNDCSNASIVYEAADLIMHTPGCVPCVREGNSSRCWV 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
               ::| :| :     |  :| |:|::| :|  |||:|||||| | | :|:|  ||: |
Db        241 ALTPTLAARNVTIPTTTIRRHVDLLVGAAAFCSAMYVGDLCGSVFLVSQLFTFSPRRHVT 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
              :|||||||||| ::|||||||||||||||  ::::  :|:|: ::|:::|||||: |||
Db        301 LQDCNCSIYPGHVSGHRMAWDMMMNWSPTTALVVSQLLRIPQAVVDMVAGAHWGVLAGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
              |:|| | ||||:::||| ||||  ||  ||: |:|| || |:|  || |||||:||||||
Db        361 YYSMAGNWAKVLIVMLLFAGVDGDTHVTGGAQAKTTNRLVSMFASGPSQKIQLINTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
              |||||||||||| |||:|:|||||||||||||||: ||:|: |  ||| : | :  ::
Db        421 HINRTALNCNDSLQTGFLAALFYTHSFNSSGCPERMAQCRTIDKFDQGWGPITYAE--SS 478

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
              | |||||||| || :| | |||||||||||||||||||| | ||| ||||||| |
Db        479 RSDQRPYCWHYPPPQCTIVPASEVCGPVYCFTPSPVVVGTTDRFGVPTYRWGENETDVLL 538

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
              ||:|||| |:|||||||:|:||||| |||      |    | ||||||||||: || |
Db        539 LNNTRPPQGNWFGCTWMNSTGFTKTCGGPPCNIGGVGN--NTLTCPTDCFRKHPEATYTK 596

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIPKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
              ||||||||||::|||||||||||||:||||:|||||||||| ||||:|||:||:|||
Db        597 CGSGPWLTPRCMVDYPYRLWHYPCTVNFTIFKVRMYVGGVEHRLNAACNWTRGERCDLED 656

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
              ||| :||||| ||||| :||||:: |||||||:|||||||||::||:  |:  : ::|
Db        657 RDRPELSPLLLSTTEWQVLPCSFTTLPALSTGLIHLHQNIVDVQYLYGIGSAVVSFAIKW 716

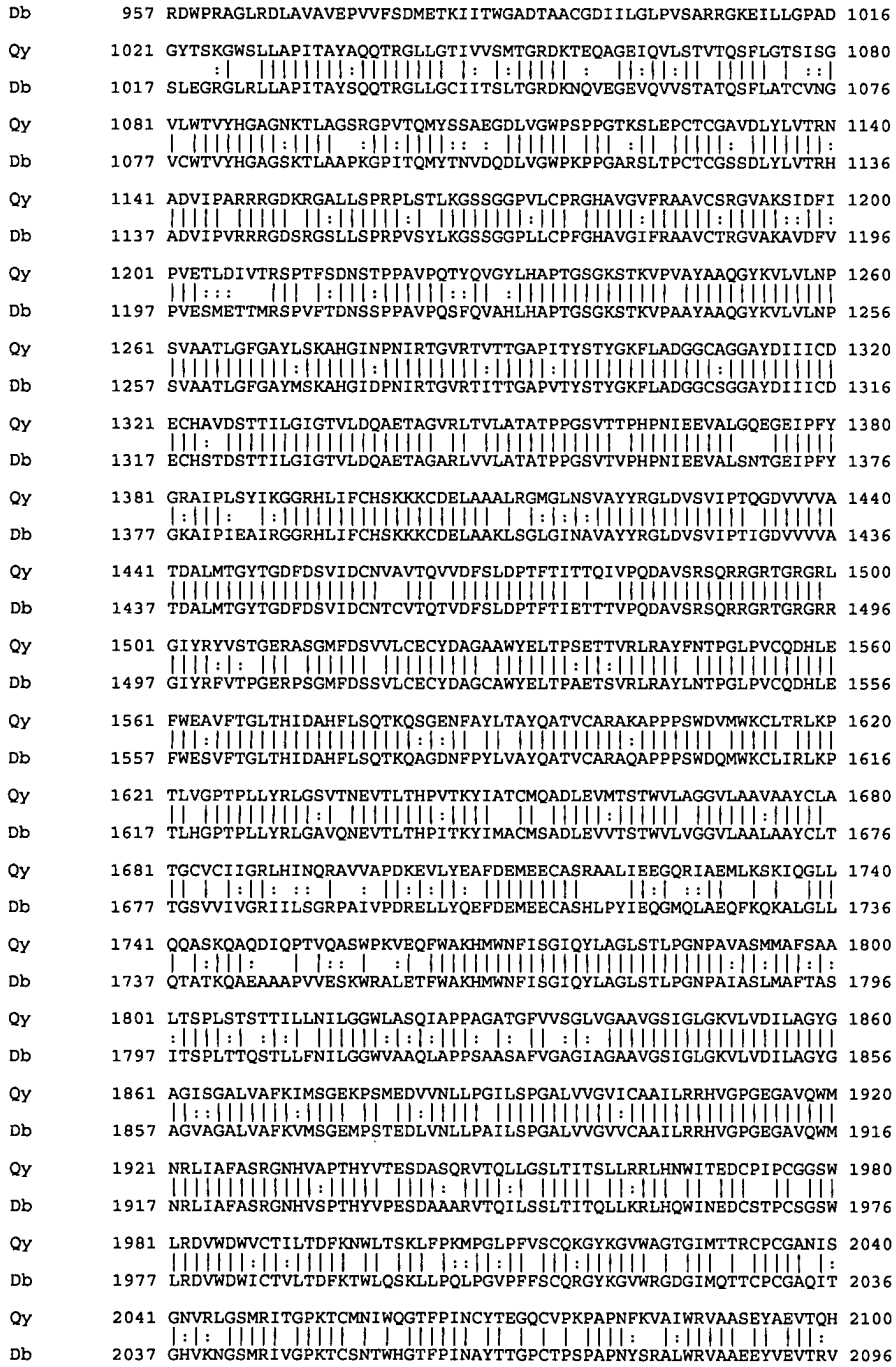
Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
              |:|:||||||||||||||::|: ||||||| ||:|::|| |  :| | |::|| |||
Db        717 EYVLLLFLLLADARVCACLWMMLLIAQAEAALENLVVLNSASVAGAHGILSFLVFFCAAW 776

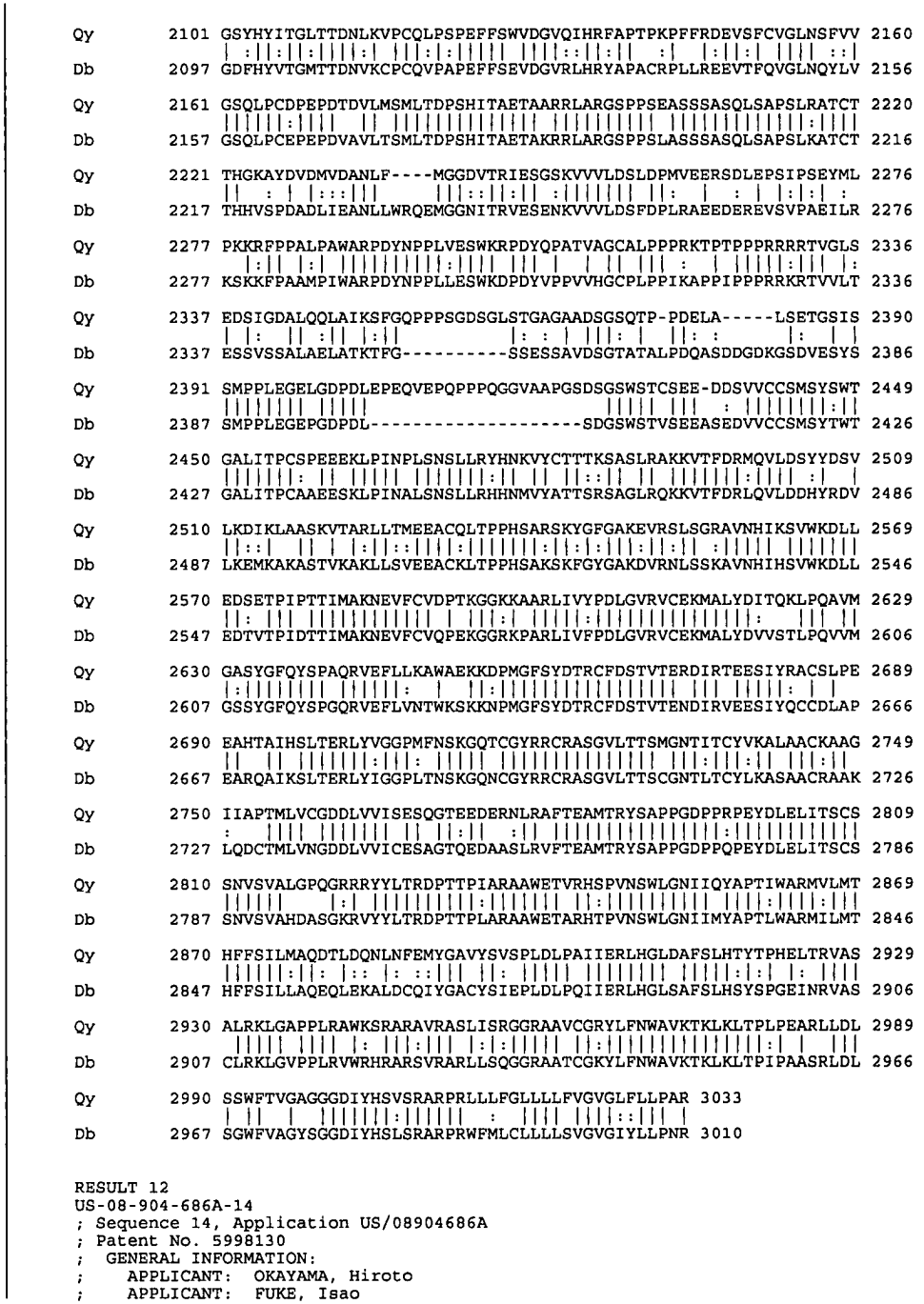
Qy        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
              |||||:|| |||:| |:|   ||||||| :||| |  :    | |: | : | ||:| ||
Db        777 YIKGRLVPGATYALYGVWPLLLLLLALPPRAYAMDREMAASCGGAVFVGLVLLTLSPYYK 836

Qy        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
               |:| :||| |  |  || :  | ||:  ||||| ||  :   :| ::|||| |:|:|
Db        837 VFLARLIWWLQYFTTRAEADLHVWIPPLNARGGRDAIILLMCAVHPELIFDITKLLIAIL 896

Qy        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
              ||  :|:  :|||||||||  |:  | : | :||| ||||| : ||  ||||||:||||:
Db        897 GPLMVLQAGITRVPYFVRAQGLIHACMLVRKVAGGHYVQMAFMKLGALTGTYIYNHLTPL 956

Qy        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
               ||  :|||||||||||::|| || |:| |||:|||||||: ||||||| |:|:||||||
```





RESULT 12
US-08-904-686A-14
; Sequence 14, Application US/08904686A
; Patent No. 5998130
; GENERAL INFORMATION:
; APPLICANT: OKAYAMA, Hiroto
; APPLICANT: FUKE, Isao

```
;    APPLICANT:  MORI, Chisato
;    APPLICANT:  TAKAMIZAWA, Akahisa
;    APPLICANT:  YOSHIDA, Iwao
;    TITLE OF INVENTION:  NON-A, NON-B HEPATITIS VIRUS GENOMIC
;    TITLE OF INVENTION:  cDNA AND ANTIGEN POLYPEPTIDE
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Armstrong, Westerman, Hattori, McLeland &
;      ADDRESSEE:  Naughton
;      STREET:  1725 K St. N.W. Suite 1000
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20006
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 in, 1.44Mb
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS, Version 5.0
;      SOFTWARE:  ASCII
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/904,686A
;      FILING DATE:  01-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/324,977
;      FILING DATE:  18-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 2-167466
;      FILING DATE:  25-JUN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 2-230921
;      FILING DATE:  31-AUG-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 2-305605
;      FILING DATE:  09-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/099,706
;      FILING DATE:  30-JUL-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/769,996
;      FILING DATE:  02-OCT-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/635,451
;      FILING DATE:  28-DEC-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McLeland, Le-Nhung
;      REGISTRATION NUMBER:  31,541
;      REFERENCE/DOCKET NUMBER:  900703G
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 659-2930
;      TELEFAX:  (202) 887-0357
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3010 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-904-686A-14

  Query Match             74.3%;  Score 11984.5;  DB 2;  Length 3010;
  Best Local Similarity   71.5%;  Pred. No. 0;
  Matches 2175;  Conservative  352;  Mismatches  472;  Indels   45;  Gaps    8;

Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRAPRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              ||||||| ||  |::| :|||||||||||||||||||||||||||||| ||| ||||:|
Db         61 RRQPIPKARRPEGRTWAQPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPTDPRRRSRNLG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
              ||||||||||||||||:|||||| ||||||||||||||||:||||||||||||||||||
Db        121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
              ||||:||| || || |:|  | |||||:| || ::    ::| |||||| : ||:|:||:
Db        181 LLSCLTTPASAYEVHNVSGIYHVTNDCSNASIVYEAADLIMHTPGCVPCVREGNSSRCWV 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
               ::| :| :      |  :| |:|::| :|  |||:|||||| | | :|:|  ||: |
Db        241 ALTPTLAARNVTIPTTTIRRHVDLLVGAAAFCSAMYVGDLCGSVFLVSQLFTFSPRRHVT 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
              :|||||||||| ::|||||||||||||||  ::::  :|:|: ::|:::|||||: |||
Db        301 LQDCNCSIYPGHVSGHRMAWDMMMNWSPTTALVVSQLLRIPQAVVDMVAGAHWGVLAGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
              |:|| | ||||:::::|| |||| || ||: |:|| || |:|  || |||||:||||||
Db        361 YYSMAGNWAKVLIVMLLFAGVDGDTHVTGGAQAKTTNRLVSMFASGPSQKIQLINTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
              |||||||||||||| |||:|:|||||||||||||||: ||:|: |  ||| : | :  ::
Db        421 HINRTALNCNDSLQTGFLAALFYTHSFNSSGCPERMAQCRTIDKFDQGWGPITYAE--SS 478

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
              | ||||||||| || :| |  |||||||||||||||||||| | ||| ||||||||| |
Db        479 RSDQRPYCWHYPPPQCTIVPASEVCGPVYCFTPSPVVVGTTDRFGVPTYRWGENETDVLL 538

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
              ||:|||| |:|||||||||:|:||||| |||      |    | ||||||||||: || |
Db        539 LNNTRPPQGNWFGCTWMNSTGFTKTCGGPPCNIGGVGN--NTLTCPTDCFRKHPEATYTK 596

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIPKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
              ||||||||||::|||||||||||||:||||:|||||||||| ||||:|||:||:|||
Db        597 CGSGPWLTPRCMVDYPYRLWHYPCTVNFTIFKVRMYVGGVEHRLNAACNWTRGERCDLED 656

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
              ||| :|||| ||||| :|||:: |||||||:|||||||||::||:  |:  : ::|
Db        657 RDRPELSPLLLSTTEWQVLPCSFTTLPALSTGLIHLHQNIVDVQYLYGIGSAVVSFAIKW 716

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
              |:|:||||||||||||||::|: ||||||| ||:|::|| |  :| | |::|| |||
Db        717 EYVLLLFLLLADARVCACLWMMLLIAQAEAALENLVVLNSASVAGAHGILSFLVFFCAAW 776

Qy        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
              |||||:|| |||:| |:|   ||||||| :||| |  :    | |: | : | ||:| ||
Db        777 YIKGRLVPGATYALYGVWPLLLLLLALPPRAYAMDREMAASCGGAVFVGLVLLTLSPYYK 836

Qy        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
              |:| :||| |  |  || :  | ||:  |||| ||  :   :| ::|||| |:|:|
Db        837 VFLARLIWWLQYFTTRAEADLHVWIPPLNARGGRDAIILLMCAVHPELIFDITKLLIAIL 896

Qy        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
              ||  :|:  :|||||||| |:  | : | :||| |||| : ||  ||||||:||||:
Db        897 GPLMVLQAGITRVPYFVRAQGLIHACMLVRKVAGGHYVQMAFMKLGALTGTYIYNHLTPL 956

Qy        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
               ||  :|||||||||||:|| || |:| |||:|||||||: |||||||  |:|:||||||
Db        957 RDWPRAGLRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGKEILLGPAD 1016

Qy       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080
                 :|  |||||||:|||||| |: |:|||| :  ||:||:|| |||| | ::|
Db       1017 SLEGRGLRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNG 1076

Qy       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
              | |||||||:|||| :||:|||||:: : |||||| ||| :|| |||||: |||||||:
Db       1077 VCWTVYHGAGSKTLAAPKGPITQMYTNVDQDLVGWPKPPGARSLTPCTCGSSDLYLVTRH 1136

Qy       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200
              ||||| ||||| ||:|||||:| |||||||:||| |||||:||||||:|||||::||:
Db       1137 ADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPFGHAVGIFRAAVCTRGVAKAVDFV 1196

Qy       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260
              |||:::   ||| |:|||:|||||::|| :|||||||||||| ||||||||||||
Db       1197 PVESMETTMRSPVFTDNSSPPAVPQSFQVAHLHAPTGSGKSTKVPAAYAAQGYKVLVLNP 1256
```

```
Qy       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320
Db       1257 SVAATLGFGAYMSKAHGIDPNIRTGVRTITTGAPVTYSTYGKFLADGGCSGGAYDIIICD 1316

Qy       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380
Db       1317 ECHSTDSTTILGIGTVLDQAETAGARLVVLATATPPGSVTVPHPNIEEVALSNTGEIPFY 1376

Qy       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440
Db       1377 GKAIPIEAIRGGRHLIFCHSKKKCDELAAKLSGLGINAVAYYRGLDVSVIPTIGDVVVVA 1436

Qy       1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500
Db       1437 TDALMTGYTGDFDSVIDCNTCVTQTVDFSLDPTFTIETTTVPQDAVSRSQRRGRTGRGRR 1496

Qy       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560
Db       1497 GIYRFVTPGERPSGMFDSSVLCECYDAGCAWYELTPAETSVRLRAYLNTPGLPVCQDHLE 1556

Qy       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
Db       1557 FWESVFTGLTHIDAHFLSQTKQAGDNFPYLVAYQATVCARAQAPPPSWDQMWKCLIRLKP 1616

Qy       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680
Db       1617 TLHGPTPLLYRLGAVQNEVTLTHPITKYIMACMSADLEVVTSTWVLVGGVLAALAAYCLT 1676

Qy       1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740
Db       1677 TGSVVIVGRIILSGRPAIVPDRELLYQEFDEMEECASHLPYIEQGMQLAEQFKQKALGLL 1736

Qy       1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
Db       1737 QTATKQAEAAAPVVESKWRALETFWAKHMWNFISGIQYLAGLSTLPGNPAIASLMAFTAS 1796

Qy       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860
Db       1797 ITSPLTTQSTLLFNILGGWVAAQLAPPSAASAFVGAGIAGAAVGSIGLGKVLVDILAGYG 1856

Qy       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920
Db       1857 AGVAGALVAFKVMSGEMPSTEDLVNLLPAILSPGALVVGVVCAAILRRHVGPGEGAVQWM 1916

Qy       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980
Db       1917 NRLIAFASRGNHVSPTHYVPESDAAARVTQILSSLTITQLLKRLHQWINEDCSTPCSGSW 1976

Qy       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040
Db       1977 LRDVWDWICTVLTDFKTWLQSKLLPQLPGVPFFSCQRGYKGVWRGDGIMQTTCPCGAQIT 2036

Qy       2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100
Db       2037 GHVKNGSMRIVGPKTCSNTWHGTFPINAYTTGPCTPSPAPNYSRALWRVAAEEYVEVTRV 2096

Qy       2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160
Db       2097 GDFHYVTGMTTDNVKCPCQVPAPEFFSEVDGVRLHRYAPACRPLLREEVTFQVGLNQYLV 2156

Qy       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220
Db       2157 GSQLPCEPEPDVAVLTSMLTDPSHITAETAKRRLARGSPPSLASSSASQLSAPSLKATCT 2216

Qy       2221 THGKAYDVDMVDANLF----MGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYML 2276
Db       2217 THHVSPDADLIEANLLWRQEMGGNITRVESENKVVVLDSFDPLRAEEDEREVSVPAEILR 2276

Qy       2277 PKKRFPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLS 2336
Db       2277 KSKKFPAAMPIWARPDYNPPLLESWKDPDYVPPVVHGCPLPPIKAPPIPPPRRKRTVVLT 2336

Qy       2337 EDSIGDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTP-PDELA-----LSETGSIS 2390
```

```
Db       2337 ESSVSSALAELATKTFG----------SSESSAVDSGTATALPDQASDDGDKGSDVESYS 2386

Qy       2391 SMPPLEGELGDPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEE-DDSVVCCSMSYSWT 2449
Db       2387 SMPPLEGEPGDPDL--------------------SDGSWSTVSEEASEDVVCCSMSYTWT 2426

Qy       2450 GALITPCSPEEEKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSV 2509
Db       2427 GALITPCAAEESKLPINALSNSLLRHHNMVYATTSRSAGLRQKKVTFDRLQVLDDHYRDV 2486

Qy       2510 LKDIKLAASKVTARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLL 2569
Db       2487 LKEMKAKASTVKAKLLSVEEACKLTPPHSAKSKFGYGAKDVRNLSSKAVNHIHSVWKDLL 2546

Qy       2570 EDSETPIPTTIMAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVM 2629
Db       2547 EDTVTPIDTTIMAKNEVFCVQPEKGGRKPARLIVFPDLGVRVCEKMALYDVVSTLPQVVM 2606

Qy       2630 GASYGFQYSPAQRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPE 2689
Db       2607 GSSYGFQYSPGQRVEFLVNTWKSKKNPMGFSYDTRCFDSTVTENDIRVEESIYQCCDLAP 2666

Qy       2690 EAHTAIHSLTERLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAG 2749
Db       2667 EARQAIKSLTERLYIGGPLTNSKGQNCGYRRCRASGVLTTSCGNTLTCYLKASAACRAAK 2726

Qy       2750 IIAPTMLVCGDDLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCS 2809
Db       2727 LQDCTMLVNGDDLVVICESAGTQEDAASLRVFTEAMTRYSAPPGDPPQPEYDLELITSCS 2786

Qy       2810 SNVSVALGPQGRRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMT 2869
Db       2787 SNVSVAHDASGKRVYYLTRDPTTPLARAAWETARHTPVNSWLGNIIMYAPTLWARMILMT 2846

Qy       2870 HFFSILMAQDTLDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVAS 2929
Db       2847 HFFSILLAQEQLEKALDCQIYGACYSIEPLDLPQIIERLHGLSAFSLHSYSPGEINRVAS 2906

Qy       2930 ALRKLGAPPLRAWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDL 2989
Db       2907 CLRKLGVPPLRVWRHRARSVRARLLSQGGRAATCGKYLFNWAVKTKLKLTPIPAASRLDL 2966

Qy       2990 SSWFTVGAGGGDIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
Db       2967 SGWFVAGYSGGDIYHSLSRARPRWFMLCLLLLSVGVGIYLLPNR 3010
```

```
RESULT 13
US-09-315-850-2
; Sequence 2, Application US/09315850
; Patent No. 6217872
;  GENERAL INFORMATION:
;    APPLICANT:  OKAYAMA, Hiroto
;    APPLICANT:  FUKE, Isao
;    APPLICANT:  MORI, Chisato
;    APPLICANT:  TAKAMIZAWA, Akahisa
;    APPLICANT:  YOSHIDA, Iwao
;    TITLE OF INVENTION:  NON-A, NON-B HEPATITIS VIRUS GENOMIC
;    TITLE OF INVENTION:  cDNA AND ANTIGEN POLYPEPTIDE
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Armstrong, Westerman, Hattori, McLeland &
;      ADDRESSEE:  Naughton
;      STREET:  1725 K St. N.W. Suite 1000
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20006
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 in, 1.44Mb
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS, Version 5.0
```

```
;      SOFTWARE:  ASCII
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/315,850
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/904,686
;      FILING DATE:  01-AUG-1997
;      APPLICATION NUMBER:  US 08/324,977
;      FILING DATE:  18-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 2-167466
;      FILING DATE:  25-JUN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 2-230921
;      FILING DATE:  31-AUG-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 2-305605
;      FILING DATE:  09-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/099,706
;      FILING DATE:  30-JUL-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/769,996
;      FILING DATE:  02-OCT-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/635,451
;      FILING DATE:  28-DEC-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McLeland, Le-Nhung
;      REGISTRATION NUMBER:  31,541
;      REFERENCE/DOCKET NUMBER:  900703G
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 659-2930
;      TELEFAX:  (202) 887-0357
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3010 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-315-850-2

  Query Match             74.3%;  Score 11984.5;  DB 3;  Length 3010;
  Best Local Similarity   71.5%;  Pred. No. 0;
  Matches 2175;  Conservative  352;  Mismatches  472;  Indels   45;  Gaps    8;

Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRAPRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              ||||||| ||  |::| :|||||||||||||||||||||||||||||| ||| ||||:|
Db         61 RRQPIPKARRPEGRTWAQPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPTDPRRRSRNLG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
              ||||||||||||||||:|||||| |||||||||||||||:|||||||||||||||||||
Db        121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
              ||||:||| || || |:|  | |||||:| || ::    ::| |||||| : ||:|:||:
Db        181 LLSCLTTPASAYEVHNVSGIYHVTNDCSNASIVYEAADLIMHTPGCVPCVREGNSSRCWV 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
               ::| :| :     |  :| |:|::| :|  |||:|||||| | | :|:|  ||: |
Db        241 ALTPTLAARNVTIPTTTIRRHVDLLVGAAAFCSAMYVGDLCGSVPLVSQLFTFSPRRHVT 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
              :|||||||||| ::|||||||||||||||  ::::  :|:|: ::|:::|||||: |||
Db        301 LQDCNCSIYPGHVSGHRMAWDMMMNWSPTTALVVSQLLRIPQAVVDMVAGAHWGVLAGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
              |:|| | ||||:::::|| |||| ||  ||: |:|| || |:|  || |||||:|||||
Db        361 YYSMAGNWAKVLIVMLLFAGVDGDTHVTGGAQAKTTNRLVSMFASGPSQKIQLINTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
              |||||||||||| |||:|:|||||||||||||||||: ||:|: |  ||| : | :  ::
Db        421 HINRTALNCNDSLQTGFLAALFYTHSFNSSGCPERMAQCRTIDKFDQGWGPITYAE--SS 478

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
              | |||||||| || :| |  ||||||||||||||||||||| | ||| ||||||||  |
Db        479 RSDQRPYCWHYPPPQCTIVPASEVCGPVYCFTPSPVVVGTTDRFGVPTYRWGENETDVLL 538

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
              ||:|||| |:|||||||:|:||||| |||      |    | ||||||||||: || |
Db        539 LNNTRPPQGNWFGCTWMNSTGFTKTCGGPPCNIGGVGN--NTLTCPTDCFRKHPEATYTK 596

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
              ||||||||||::|||||||||||||:||||:|||||||||| ||||:|||:||:|||
Db        597 CGSGPWLTPRCMVDYPYRLWHYPCTVNFTIFKVRMYVGGVEHRLNAACNWTRGERCDLED 656

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
              ||| :||||| ||||| :||||:: ||||||||:|||||||||::||:  |:  : ::|
Db        657 RDRPELSPLLLSTTEWQVLPCSFTTLPALSTGLIHLHQNIVDVQYLYGIGSAVVSFAIKW 716

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
              |:|:|||||||||||||||::|: ||||||| ||:|::|| |  :| | |::|| |||
Db        717 EYVLLLFLLLADARVCACLWMMLLIAQAEAALENLVVLNSASVAGAHGILSFLVFFCAAW 776

Qy        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
              |||||:|| |||:| |:|   ||||||| :||| |  :    | |: | : | ||:| ||
Db        777 YIKGRLVPGATYALYGVWPLLLLLLALPPRAYAMDREMAASCGGAVFVGLVLLTLSPYYK 836

Qy        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
              |:| :||| |  |  || :  | ||:  ||||| ||  :   :| ::|||||  |:|:|
Db        837 VFLARLIWWLQYFTTRAEADLHVWIPPLNARGGRDAIILLMCAVHPELIFDITKLLIAIL 896

Qy        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
              ||  :|:  :|||||||||  |:  | : | :||| ||||| : ||  ||||||:||||:
Db        897 GPLMVLQAGITRVPYFVRAQGLIHACMLVRKVAGGHYVQMAFMKLGALTGTYIYNHLTPL 956

Qy        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
              ||  :|||||||||::|| || |:|  |||:|||||||: ||||||| |:|:||||||
Db        957 RDWPRAGLRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGKEILLGPAD 1016

Qy       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080
                :|  ||||||||:|||||||| |: |:|||| :  ||:||:|| ||||| | ::|
Db       1017 SLEGRGLRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNG 1076

Qy       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
              | ||||||:|||  :||:|||:: : |||||| ||| :|| |||||: ||||||:
Db       1077 VCWTVYHGAGSKTLAAPKGPITQMYTNVDQDLVGWPKPPGARSLTPCTCGSSDLYLVTRH 1136

Qy       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDPI 1200
              ||||| |||||  ||:|||||:| |||||||:||| |||||:|||||:|||||::||:
Db       1137 ADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPFGHAVGIFRAAVCTRGVAKAVDFV 1196

Qy       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260
              |||:::   ||| |:|||:||||||::|| :|||||||||||| ||||||||||||
Db       1197 PVESMETTMRSPVFTDNSSPPAVPQSFQVAHLHAPTGSGKSTKVPAAYAAQGYKVLVLNP 1256

Qy       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320
              |||||||||||:||||||:|||||||||:|||||:|||||||||||||:||||||||||||
Db       1257 SVAATLGFGAYMSKAHGIDPNIRTGVRTITTGAPVTYSTYGKFLADGGCSGGAYDIIICD 1316

Qy       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380
              |||: |||||||||||||||| || |||||||||||  ||||||||||  ||||||
Db       1317 ECHSTDSTTILGIGTVLDQAETAGARLVVLATATPPGSVTVPHPNIEEVALSNTGEIPFY 1376

Qy       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440
              |:|||:  |:|||||||||||||||||| | |:|:|:|||||||||||||| |||||||
Db       1377 GKAIPIEAIRGGRHLIFCHSKKKCDELAAKLSGLGINAVAYYRGLDVSVIPTIGDVVVVA 1436

Qy       1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500
              |||||||||||||||||| ||| |||||||||| |  ||||||||||||||||||
Db       1437 TDALMTGYTGDFDSVIDCNTCVTQTVDFSLDPTFTIETTTVPQDAVSRSQRRGRTGRGRR 1496
```

```
Qy       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560
Db       1497 GIYRFVTPGERPSGMFDSSVLCECYDAGCAWYELTPAETSVRLRAYLNTPGLPVCQDHLE 1556

Qy       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
Db       1557 FWESVFTGLTHIDAHFLSQTKQAGDNFPYLVAYQATVCARAQAPPPSWDQMWKCLIRLKP 1616

Qy       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680
Db       1617 TLHGPTPLLYRLGAVQNEVTLTHPITKYIMACMSADLEVVTSTWVLVGGVLAALAAYCLT 1676

Qy       1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740
Db       1677 TGSVVIVGRIILSGRPAIVPDRELLYQEFDEMEECASHLPYIEQGMQLAEQFKQKALGLL 1736

Qy       1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
Db       1737 QTATKQAEAAAPVVESKWRALETFWAKHMWNFISGIQYLAGLSTLPGNPAIASLMAFTAS 1796

Qy       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860
Db       1797 ITSPLTTQSTLLFNILGGWVAAQLAPPSAASAFVGAGIAGAAVGSIGLGKVLVDILAGYG 1856

Qy       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920
Db       1857 AGVAGALVAFKVMSGEMPSTEDLVNLLPAILSPGALVVGVVCAAILRRHVGPGEGAVQWM 1916

Qy       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980
Db       1917 NRLIAFASRGNHVSPTHYVPESDAAARVTQILSSLTITQLLKRLHQWINEDCSTPCSGSW 1976

Qy       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040
Db       1977 LRDVWDWICTVLTDFKTWLQSKLLPQLPGVPFFSCQRGYKGVWRGDGIMQTTCPCGAQIT 2036

Qy       2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100
Db       2037 GHVKNGSMRIVGPKTCSNTWHGTFPINAYTTGPCTPSPAPNYSRALWRVAAEEYVEVTRV 2096

Qy       2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160
Db       2097 GDFHYVTGMTTDNVKCPCQVPAPEFFSEVDGVRLHRYAPACRPLLREEVTFQVGLNQYLV 2156

Qy       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220
Db       2157 GSQLPCEPEPDVAVLTSMLTDPSHITAETAKRRLARGSPPSLASSSASQLSAPSLKATCT 2216

Qy       2221 THGKAYDVDMVDANLF----MGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYML 2276
Db       2217 THHVSPDADLIEANLLWRQEMGGNITRVESENKVVVLDSFDPLRAEEDEREVSVPAEILR 2276

Qy       2277 PKKRFPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLS 2336
Db       2277 KSKKFPAAMPIWARPDYNPPLLESWKDPDYVPPVVHGCPLPPIKAPPIPPPRRKRTVVLT 2336

Qy       2337 EDSIGDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTP-PDELA-----LSETGSIS 2390
Db       2337 ESSVSSALAELATKTFG----------SSESSAVDSGTATALPDQASDDGDKGSDVESYS 2386

Qy       2391 SMPPLEGELGDPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEE-DDSVVCCSMSYSWT 2449
Db       2387 SMPPLEGEPGDPDL-------------------SDGSWSTVSEEASEDVVCCSMSYTWT 2426

Qy       2450 GALITPCSPEEEKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSV 2509
Db       2427 GALITPCAAEESKLPINALSNSLLRHHNMVYATTSRSAGLRQKKVTFDRLQVLDDHYRDV 2486

Qy       2510 LKDIKLAASKVTARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLL 2569
Db       2487 LKEMKAKASTVKAKLLSVEEACKLTPPHSAKSKFGYGAKDVRNLSSKAVNHIHSVWKDLL 2546

Qy       2570 EDSETPIPTTIMAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVM 2629

Db       2547 EDTVTPIDTTIMAKNEVFCVQPEKGGRKPARLIVFPDLGVRVCEKMALYDVVSTLPQVVM 2606

Qy       2630 GASYGFQYSPAQRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPE 2689
Db       2607 GSSYGFQYSPGQRVEFLVNTWKSKKNPMGFSYDTRCFDSTVTENDIRVEESIYQCCDLAP 2666

Qy       2690 EAHTAIHSLTERLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAG 2749
Db       2667 EARQAIKSLTERLYIGGPLTNSKGQNCGYRRCRASGVLTTSCGNTLTCYLKASAACRAAK 2726

Qy       2750 IIAPTMLVCGDDLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCS 2809
Db       2727 LQDCTMLVNGDDLVVICESAGTQEDAASLRVFTEAMTRYSAPPGDPPQPEYDLELITSCS 2786

Qy       2810 SNVSVALGPQGRRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMT 2869
Db       2787 SNVSVAHDASGKRVYYLTRDPTTPLARAAWETARHTPVNSWLGNIIMYAPTLWARMILMT 2846

Qy       2870 HFFSILMAQDTLDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVAS 2929
Db       2847 HFFSILLAQEQLEKALDCQIYGACYSIEPLDLPQIIERLHGLSAFSLHSYSPGEINRVAS 2906

Qy       2930 ALRKLGAPPLRAWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDL 2989
Db       2907 CLRKLGVPPLRVWRRHRARSVRARLLSQGGRAATCGKYLFNWAVKTKLKLTPIPAASRLDL 2966

Qy       2990 SSWFTVGAGGGDIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
Db       2967 SGWFVAGYSGGDIYHSLSRARPRWFMLCLLLLSVGVGIYLLPNR 3010
```

RESULT 14
US-09-315-850-14
; Sequence 14, Application US/09315850
; Patent No. 6217872
; GENERAL INFORMATION:
; APPLICANT: OKAYAMA, Hiroto
; APPLICANT: FUKE, Isao
; APPLICANT: MORI, Chisato
; APPLICANT: TAKAMIZAWA, Akahisa
; APPLICANT: YOSHIDA, Iwao
; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
; TITLE OF INVENTION: cDNA AND ANTIGEN POLYPEPTIDE
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Armstrong, Westerman, Hattori, McLeland &
; ADDRESSEE: Naughton
; STREET: 1725 K St. N.W. Suite 1000
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 in, 1.44Mb
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/315,850
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/904,686
; FILING DATE: 01-AUG-1997
; APPLICATION NUMBER: US 08/324,977
; FILING DATE: 18-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-167466
; FILING DATE: 25-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-230921
; FILING DATE: 31-AUG-1990
; PRIOR APPLICATION DATA:

```
;      APPLICATION NUMBER:  JP 2-305605
;      FILING DATE:  09-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/099,706
;      FILING DATE:  30-JUL-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/769,996
;      FILING DATE:  02-OCT-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/635,451
;      FILING DATE:  28-DEC-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McLeland, Le-Nhung
;      REGISTRATION NUMBER:  31,541
;      REFERENCE/DOCKET NUMBER:  900703G
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 659-2930
;      TELEFAX:  (202) 887-0357
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3010 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-315-850-14

  Query Match             74.3%;  Score 11984.5;  DB 3;  Length 3010;
  Best Local Similarity   71.5%;  Pred. No. 0;
  Matches 2175;  Conservative  352;  Mismatches  472;  Indels   45;  Gaps    8;

Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRAPRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
              ||||||| ||  |::| :|||||||||||||||||||||||||||||| ||| ||||:|
Db         61 RRQPIPKARRPEGRTWAQPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPTDPRRRSRNLG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
              |||||||||||||||||:|||||||  ||||||||||||||||:||||||||||||||||
Db        121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
              ||||:||| || || |:|  | |||||:| || ::    ::| |||||| : ||:|:||:
Db        181 LLSCLTTPASAYEVHNVSGIYHVTNDCSNASIVYEAADLIMHTPGCVPCVREGNSSRCWV 240

Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
               ::| :| :      | :| |:|::| :|  |||:||||||| | | :|:|  ||: |
Db        241 ALTPTLAARNVTIPTTTIRRHVDLLVGAAAFCSAMYVGDLCGSVFLVSQLFTFSPRRHVT 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
              :|||||||||| ::|||||||||||||||  ::::  :|:|: ::|:::|||||: |||
Db        301 LQDCNCSIYPGHVSGHRMAWDMMMNWSPTTALVVSQLLRIPQAVVDMVAGAHWGVLAGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
              |:|| | ||||:::||  ||||   ||  ||: |:|| || |:|  || |||||:||||||
Db        361 YYSMAGNWAKVLIVMLLFAGVDGDTHVTGGAQAKTTNRLVSMFASGPSQKIQLINTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
              |||||||||||| |||:|:|||||||||||||||: ||:|: |  ||| : | :  ::
Db        421 HINRTALNCNDSLQTGFLAALFYTHSFNSSGCPERMAQCRTIDKFDQGWGPITYAE--SS 478

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
              | |||||||| || :| |  ||||||||||||||||||| | ||| ||||||| |
Db        479 RSDQRPYCWHYPPPQCTIVPASEVCGPVYCFTPSPVVVGTTDRFGVPTYRWGENETDVLL 538

Qy        541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
              ||:|||| |:|||||||:|:||||| |||       |    | ||||||||||: || |
Db        539 LNNTRPPQGNWFGCTWMNSTGFTKTCGGPPCNIGGVGN--NTLTCPTDCFRKHPEATYTK 596

Qy        601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
              ||||||||||::|||||||||||||:||||:|||||||||| ||||:|||:||:|||
Db        597 CGSGPWLTPRCMVDYPYRLWHYPCTVNFTIFKVRMYVGGVEHRLNAACNWTRGERCDLED 656

Qy        661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
              ||| :||||| ||||| :||||:: ||||||||:|||||||||::||:  |:  : ::|
Db        657 RDRPELSPLLLSTTEWQVLPCSFTTLPALSTGLIHLHQNIVDVQYLYGIGSAVVSFAIKW 716

Qy        721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
              |:|:||||||||||||||||::|: |||||| ||:|::|| |  :| | |::|| |||
Db        717 EYVLLLFLLLADARVCACLWMMLLIAQAEAALENLVVLNSASVAGAHGILSFLVFFCAAW 776

Qy        781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
              |||||:|| |||:| |:|   ||||||| :||| |  :    | |: | : | ||:| ||
Db        777 YIKGRLVPGATYALYGVWPLLLLLLALPPRAYAMDREMAASCGGAVFVGLVLLTLSPYYK 836

Qy        841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
              |:| :||| |  |  || :  | ||:  ||||| ||  :   :| ::||||| |:|:|
Db        837 VFLARLIWWLQYFTTRAEADLHVWIPPLNARGGRDAIILLMCAVHPELIFDITKLLIAIL 896

Qy        901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
              ||  :|:  :|||||||||  |:  | : | :||| ||||| : ||  ||||||:||||:
Db        897 GPLMVLQAGITRVPYFVRAQGLIHACMLVRKVAGGHYVQMAFMKLGALTGTYIYNHLTPL 956

Qy        961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
              ||  :|||||||||||::|| || |:| |||:|||||||: ||||||| |:|:|||||
Db        957 RDWPRAGLRDLAVAVEPVVFSDMETKIITWGADTAACGDIILGLPVSARRGKEILLGPAD 1016

Qy       1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080
                :|  ||||||||:|||||||| |: |:|||| :  ||:||:|| ||||| | ::|
Db       1017 SLEGRGLRLLAPITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTATQSFLATCVNG 1076

Qy       1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
              | ||||||||:|||| :||:||||:: : |||||| ||| :|| |||||: |||||||:
Db       1077 VCWTVYHGAGSKTLAAPKGPITQMYTNVDQDLVGWPKPPGARSLTPCTCGSSDLYLVTRH 1136

Qy       1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200
              ||||| ||||| ||:|||||:| |||||||:||| |||||:||||||:||||::||:
Db       1137 ADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPFGHAVGIFRAAVCTRGVAKAVDFV 1196

Qy       1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260
              |||:::   ||| |:|||:||||||::|| :|||||||||||| ||||||||||||
Db       1197 PVESMETTMRSPVFTDNSSPPAVPQSFQVAHLHAPTGSGKSTKVPAAYAAQGYKVLVLNP 1256

Qy       1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320
              ||||||||||:||||||:|||||||||:|||||:|||||||||||||:||||||||||
Db       1257 SVAATLGFGAYMSKAHGIDPNIRTGVRTITTGAPVTYSTYGKFLADGGCSGGAYDIIICD 1316

Qy       1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380
              |||: |||||||||||||||||| || |||||||||||| ||||||||||   ||||||
Db       1317 ECHSTDSTTILGIGTVLDQAETAGARLVVLATATPPGSVTVPHPNIEEVALSNTGEIPFY 1376

Qy       1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440
              |:|||:  |:|||||||||||||||| | |:|:|:|||||||||||| ||||||
Db       1377 GKAIPIEAIRGGRHLIFCHSKKKCDELAAKLSGLGINAVAYYRGLDVSVIPTIGDVVVVA 1436

Qy       1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500
              |||||||||||||||||  ||| |||||||||| |  |||||||||||||||||||
Db       1437 TDALMTGYTGDFDSVIDCNTCVTQTVDFSLDPTFTIETTTVPQDAVSRSQRRGRTGRGRR 1496

Qy       1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560
              ||||:|: ||| |||||| |||||||| |||||||:||:|||||| ||||||||||||
Db       1497 GIYRFVTPGERPSGMFDSSVLCECYDAGCAWYELTPAETSVRLRAYLNTPGLPVCQDHLE 1556

Qy       1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
              |||:||||||||||||||:|:|| || |||||||||:||||| ||||| ||||
Db       1557 FWESVFTGLTHIDAHFLSQTKQAGDNFPYLVAYQATVCARAQAPPPSWDQMWKCLIRLKP 1616

Qy       1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680
              || |||||||||:| ||||||||:|||| || |||||:|||||| ||||||:|||||
Db       1617 TLHGPTPLLYRLGAVQNEVTLTHPITKYIMACMSADLEVVTSTWVLVGGVLAALAAYCLT 1676

Qy       1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740
              || | |:||: :: |  : ||:|:||: |||||||||     ||:| ::||  | |  |||
Db       1677 TGSVVIVGRIILSGRPAIVPDRELLYQEFDEMEECASHLPYIEQGMQLAEQFKQKALGLL 1736
```

```
Qy       1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
Db       1737 QTATKQAEAAAPVVESKWRALETFWAKHMWNFISGIQYLAGLSTLPGNPAIASLMAFTAS 1796

Qy       1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860
Db       1797 ITSPLTTQSTLLFNILGGWVAAQLAPPSAASAFVGAGIAGAAVGSIGLGKVLVDILAGYG 1856

Qy       1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920
Db       1857 AGVAGALVAFKVMSGEMPSTEDLVNLLPAILSPGALVVGVVCAAILRRHVGPGEGAVQWM 1916

Qy       1921 NRLIAFASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980
Db       1917 NRLIAFASRGNHVSPTHYVPESDAAARVTQILSSLTITQLLKRLHQWINEDCSTPCSGSW 1976

Qy       1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040
Db       1977 LRDVWDWICTVLTDFKTWLQSKLLPQLPGVPFFSCQRGYKGVWRGDGIMQTTCPCGAQIT 2036

Qy       2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100
Db       2037 GHVKNGSMRIVGPKTCSNTWHGTFPINAYTTGPCTPSPAPNYSRALWRVAAEEYVEVTRV 2096

Qy       2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160
Db       2097 GDFHYVTGMTTDNVKCPCQVPAPEFFSEVDGVRLHRYAPACRPLLREEVTFQVGLNQYLV 2156

Qy       2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220
Db       2157 GSQLPCEPEPDVAVLTSMLTDPSHITAETAKRRLARGSPPSLASSSASQLSAPSLKATCT 2216

Qy       2221 THGKAYDVDMVDANLF----MGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYML 2276
Db       2217 THHVSPDADLIEANLLWRQEMGGNITRVESENKVVVLDSFDPLRAEEDEREVSVPAEILR 2276

Qy       2277 PKKRFPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLS 2336
Db       2277 KSKKFPAAMPIWARPDYNPPLLESWKDPDYVPPVVHGCPLPPIKAPPIPPPRRKRTVVLT 2336

Qy       2337 EDSIGDALQQLAIKSFGQPPPSGDSGLSTGAGAADSGSQTP-PDELA-----LSETGSIS 2390
Db       2337 ESSVSSALAELATKTFG----------SSESSAVDSGTATALPDQASDDGDKGSDVESYS 2386

Qy       2391 SMPPLEGELGDPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEE-DDSVVCCSMSYSWT 2449
Db       2387 SMPPLEGEPGDPDL--------------------SDGSWSTVSEEASEDVVCCSMSYTWT 2426

Qy       2450 GALITPCSPEEEKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSV 2509
Db       2427 GALITPCAAEESKLPINALSNSLLRHHNMVYATTSRSAGLRQKKVTFDRLQVLDDHYRDV 2486

Qy       2510 LKDIKLAASKVTARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLL 2569
Db       2487 LKEMKAKASTVKAKLLSVEEACKLTPPHSAKSKFGYGAKDVRNLSSKAVNHIHSVWKDLL 2546

Qy       2570 EDSETPIPTTIMAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVM 2629
Db       2547 EDTVTPIDTTIMAKNEVFCVQPEKGGRKPARLIVFPDLGVRVCEKMALYDVVSTLPQVVM 2606

Qy       2630 GASYGFQYSPAQRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPE 2689
Db       2607 GSSYGFQYSPGQRVEFLVNTWKSKKNPMGFSYDTRCFDSTVTENDIRVEESIYQCCDLAP 2666

Qy       2690 EAHTAIHSLTERLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAG 2749
Db       2667 EARQAIKSLTERLYIGGPLTNSKGQNCGYRRCRASGVLTTSCGNTLTCYLKASAACRAAK 2726

Qy       2750 IIAPTMLVCGDDLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCS 2809
Db       2727 LQDCTMLVNGDDLVVICESAGTQEDAASLRVFTEAMTRYSAPPGDPPQPEYDLELITSCS 2786

Qy       2810 SNVSVALGPQGRRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMT 2869
Db       2787 SNVSVAHDASGKRVYYLTRDPTTPLARAAWETARHTPVNSWLGNIIMYAPTLWARMILMT 2846

Qy       2870 HFFSILMAQDTLDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVAS 2929
Db       2847 HFFSILLAQEQLEKALDCQIYGACYSIEPLDLPQIIERLHGLSAFSLHSYSPGEINRVAS 2906

Qy       2930 ALRKLGAPPLRAWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDL 2989
Db       2907 CLRKLGVPPLRVWRHRARSVRARLLSQGGRAATCGKYLFNWAVKTKLKLTPIPAASRLDL 2966

Qy       2990 SSWFTVGAGGGDIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
Db       2967 SGWFVAGYSGGDIYHSLSRARPRWFMLCLLLLSVGVGIYLLPNR 3010
```

RESULT 15
US-09-539-601-33
; Sequence 33, Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
; APPLICANT: Bartenschlager, Ralf FW
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
; FILE REFERENCE: all sequences
; CURRENT APPLICATION NUMBER: US/09/539,601C
; CURRENT FILING DATE: 2001-08-30
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
; EARLIER FILING DATE: 1999-04-03
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
; LENGTH: 3010
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-539-601-33

Query Match 74.3%; Score 11984.5; DB 4; Length 3010;
Best Local Similarity 71.5%; Pred. No. 0;
Matches 2174; Conservative 353; Mismatches 477; Indels 35; Gaps 7;

```
Qy          1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
Db          1 MGTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Qy         61 RRQPIPKDRRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNVG 120
Db         61 RRQPIPKARQPEGRAWAQPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPTDPRRRSRNLG 120

Qy        121 KVIDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNFATGNLPGCSFSIFLLA 180
Db        121 KVIDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLA 180

Qy        181 LLSCITTPVSAAEVKNISTGYMVTNDCTNDSITWQLQAAVLHVPGCVPCEKVGNASQCWI 240
Db        181 LLSCLTIPASAYEVRNVSGVYHVTNDCSNASIVYEAADMIMHTPGCVPCVRENNSSRCWV 240

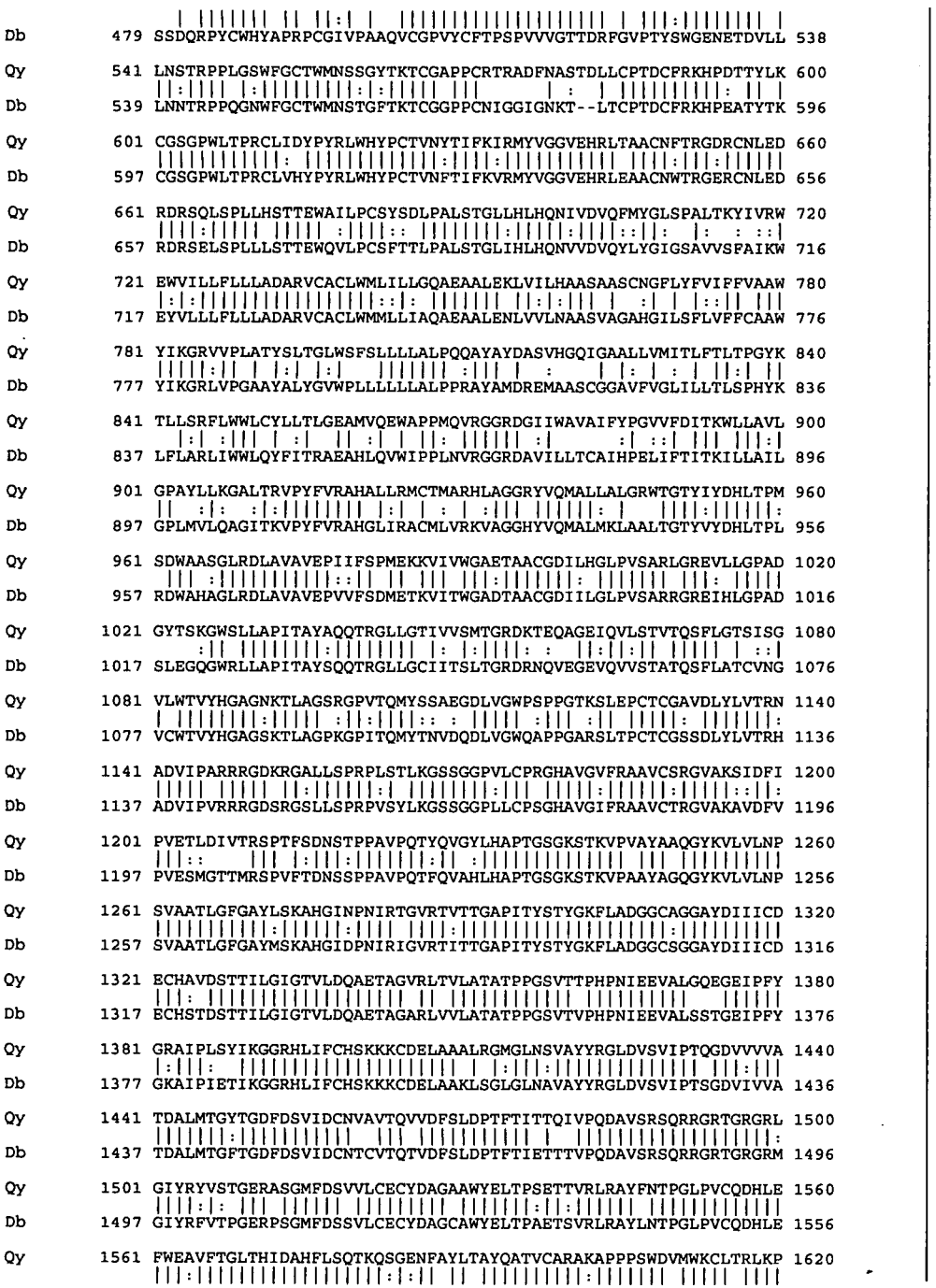
Qy        241 PVSPNVAVQRPGALTQGLRTHIDMVVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHHWF 300
Db        241 ALTPTLAARNASVPTTTIRRHVDLLVGAAALCSAMYVGDLCGSVFLVAQLFTFSPRRHET 300

Qy        301 VQDCNCSIYPGTITGHRMAWDMMMNWSPTATMILAYAMRVPEVIIDIISGAHWGVMFGLA 360
Db        301 VQDCNCSIYPGHVTGHRMAWDMMMNWSPTAALVVSQLLRIPQAVVDMVAGAHWGVLAGLA 360

Qy        361 YFSMQGAWAKVVVILLLAAGVDARTHTVGGSAAQTTGRLTSLFDMGPRQKIQLVNTNGSW 420
Db        361 YYSMVGNWAKVLIVMLLFAGVDGGTYVTGGTMAKNTLGITSLFSPGSSQKIQLVNTNGSW 420

Qy        421 HINRTALNCNDSLHTGFIASLFYTHSFNSSGCPERMSACRSIEAFRVGWGALQYEDNVTN 480
Db        421 HINRTALNCNDSLNTGFLAALFYVHKFNSSGCPERMASCSPIDAFAQGWGPITY--NESH 478

Qy        481 PEDMRPYCWHYPPRQCGVVSAKTVCGPVYCFTPSPVVVGTTDRLGAPTYTWGENETDVFL 540
```


Db 479 SSDQRPYCWHYAPRPCGIVPAAQVCGPVYCFTPSPVVVGTTDRFGVPTYSWGENETDVLL 538
Qy 541 LNSTRPPLGSWFGCTWMNSSGYTKTCGAPPCRTRADFNASTDLLCPTDCFRKHPDTTYLK 600
Db 539 LNNTRPPQGNWFGCTWMNSTGFTKTCGGPPCNIGGIGNKT--LTCPTDCFRKHPEATYTK 596
Qy 601 CGSGPWLTPRCLIDYPYRLWHYPCTVNYTIFKIRMYVGGVEHRLTAACNFTRGDRCNLED 660
Db 597 CGSGPWLTPRCLVHYPYRLWHYPCTVNFTIFKVRMYVGGVEHRLEAACNWTRGERCNLED 656
Qy 661 RDRSQLSPLLHSTTEWAILPCSYSDLPALSTGLLHLHQNIVDVQFMYGLSPALTKYIVRW 720
Db 657 RDRSELSPLLLSTTEWQVLPCSFTTLPALSTGLIHLHQNVVDVQYLYGIGSAVVSFAIKW 716
Qy 721 EWVILLFLLLADARVCACLWMLILLGQAEAALEKLVILHAASAASCNGFLYFVIFFVAAW 780
Db 717 EYVLLLFLLLADARVCACLWMMLLIAQAEAALENLVVLNAASVAGAHGILSFLVFFCAAW 776
Qy 781 YIKGRVVPLATYSLTGLWSFSLLLLALPQQAYAYDASVHGQIGAALLVMITLFTLTPGYK 840
Db 777 YIKGRLVPGAAYALYGVWPLLLLLLALPPRAYAMDREMAASCGGAVFVGLILLTLSPHYK 836
Qy 841 TLLSRFLWWLCYLLTLGEAMVQEWAPPMQVRGGRDGIIWAVAIFYPGVVFDITKWLLAVL 900
Db 837 LFLARLIWWLQYFITRAEAHLQVWIPPLNVRGGRDAVILLTCAIHPELIFTITKILLAIL 896
Qy 901 GPAYLLKGALTRVPYFVRAHALLRMCTMARHLAGGRYVQMALLALGRWTGTYIYDHLTPM 960
Db 897 GPLMVLQAGITKVPYFVRAHGLIRACMLVRKVAGGHYVQMALMKLAALTGTYVYDHLTPL 956
Qy 961 SDWAASGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDILHGLPVSARLGREVLLGPAD 1020
Db 957 RDWAHAGLRDLAVAVEPVVFSDMETKVITWGADTAACGDIILGLPVSARRGREIHLGPAD 1016
Qy 1021 GYTSKGWSLLAPITAYAQQTRGLLGTIVVSMTGRDKTEQAGEIQVLSTVTQSFLGTSISG 1080
Db 1017 SLEGQGWRLLAPITAYSQQTRGLLGCIITSLTGRDRNQVEGEVQVVSTATQSFLATCVNG 1076
Qy 1081 VLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEPCTCGAVDLYLVTRN 1140
Db 1077 VCWTVYHGAGSKTLAGPKGPITQMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYLVTRH 1136
Qy 1141 ADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFI 1200
Db 1137 ADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFV 1196
Qy 1201 PVETLDIVTRSPTFSDNSTPPAVPQTYQVGYLHAPTGSGKSTKVPVAYAAQGYKVLVLNP 1260
Db 1197 PVESMGTTMRSPVFTDNSSPPAVPQTFQVAHLHAPTGSGKSTKVPAAYAGQGYKVLVLNP 1256
Qy 1261 SVAATLGFGAYLSKAHGINPNIRTGVRTVTTGAPITYSTYGKFLADGGCAGGAYDIIICD 1320
Db 1257 SVAATLGFGAYMSKAHGIDPNIRIGVRTITTGAPITYSTYGKFLADGGCSGGAYDIIICD 1316
Qy 1321 ECHAVDSTTILGIGTVLDQAETAGVRLTVLATATPPGSVTTPHPNIEEVALGQEGEIPFY 1380
Db 1317 ECHSTDSTTILGIGTVLDQAETAGARLVVLATATPPGSVTVPHPNIEEVALSSTGEIPFY 1376
Qy 1381 GRAIPLSYIKGGRHLIFCHSKKKCDELAAALRGMGLNSVAYYRGLDVSVIPTQGDVVVVA 1440
Db 1377 GKAIPIETIKGGRHLIFCHSKKKCDELAAKLSGLGLNAVAYYRGLDVSVIPTSGDVIVVA 1436
Qy 1441 TDALMTGYTGDFDSVIDCNVAVTQVVDFSLDPTFTITTQIVPQDAVSRSQRRGRTGRGRL 1500
Db 1437 TDALMTGFTGDFDSVIDCNTCVTQTVDFSLDPTFTIETTTVPQDAVSRSQRRGRTGRGRM 1496
Qy 1501 GIYRYVSTGERASGMFDSVVLCECYDAGAAWYELTPSETTVRLRAYFNTPGLPVCQDHLE 1560
Db 1497 GIYRFVTPGERPSGMFDSSVLCECYDAGCAWYELTPAETSVRLRAYLNTPGLPVCQDHLE 1556
Qy 1561 FWEAVFTGLTHIDAHFLSQTKQSGENFAYLTAYQATVCARAKAPPPSWDVMWKCLTRLKP 1620
Db 1557 FWESVFTGLTHIDAHFLSQTKQAGDNFPYLVAYQATVCARAQAPPPSWDQMWKCLIRLKP 1616
Qy 1621 TLVGPTPLLYRLGSVTNEVTLTHPVTKYIATCMQADLEVMTSTWVLAGGVLAAVAAYCLA 1680
Db 1617 TLHGPTPLLYRLGAVQNEVTTTHPITKYIMACMSADLEVVTSTWVLVGGVLAALAAYCLT 1676
Qy 1681 TGCVCIIGRLHINQRAVVAPDKEVLYEAFDEMEECASRAALIEEGQRIAEMLKSKIQGLL 1740
Db 1677 TGSVVIVGRIILSGKPAIIPDREVLYREFDEMEECASHLPYIEQGMQLAEQFKQKAIGLL 1736
Qy 1741 QQASKQAQDIQPTVQASWPKVEQFWAKHMWNFISGIQYLAGLSTLPGNPAVASMMAFSAA 1800
Db 1737 QTATKQAEAAAPVVESKWRTIEAFWAKHMWNFISGIQYLAGLSTLPGNPAIASLMAFTAS 1796
Qy 1801 LTSPLSTSTTILLNILGGWLASQIAPPAGATGFVVSGLVGAAVGSIGLGKVLVDILAGYG 1860
Db 1797 ITSPLTTQHTLLFNILGGWVAAQLAPPSAASAFVGAGIAGAAVGSIGLGKVLVDILAGYG 1856
Qy 1861 AGISGALVAFKIMSGEKPSMEDVVNLLPGILSPGALVVGVICAAILRRHVGPGEGAVQWM 1920
Db 1857 AGVAGALVAFKVMSGELPSTEDLVNLLPAILSPGALVVGVVCAAILRRHVGPGEGAVQWM 1916
Qy 1921 NRLIAPASRGNHVAPTHYVTESDASQRVTQLLGSLTITSLLRRLHNWITEDCPIPCGGSW 1980
Db 1917 NRLIAPASRGNHVSPTHYVPESDAAARVTQILSSLTITQLLKRLHQWINEDCSTPCSGSW 1976
Qy 1981 LRDVWDWVCTILTDFKNWLTSKLFPKMPGLPFVSCQKGYKGVWAGTGIMTTRCPCGANIS 2040
Db 1977 LRDVWDWICTVLTDFKTWLQSKLLPRLPGVPFFSCQRGYKGVWRGDGIMQTTCPCGAQIT 2036
Qy 2041 GNVRLGSMRITGPKTCMNIWQGTFPINCYTEGQCVPKPAPNFKVAIWRVAASEYAEVTQH 2100
Db 2037 GHVKNGSMRIVGPRTCSNTWHGTFPINAYTTGPCTPSPAPNYSRALWRVGAEEYVEVTRV 2096
Qy 2101 GSYHYITGLTTDNLKVPCQLPSPEFFSWVDGVQIHRFAPTPKPFFRDEVSFCVGLNSFVV 2160
Db 2097 GDFHYVTGMTTDNVKCPCQVPAPEFFTEVDGVRLHRYAPACKPLLREEVTFLVGLNQYLV 2156
Qy 2161 GSQLPCDPEPDTDVLMSMLTDPSHITAETAARRLARGSPPSEASSSASQLSAPSLRATCT 2220
Db 2157 GSQLPCEPEPDVAVLTSMLTDPSHITAETAKRRLARGSPPPLASSSASQLSAPSLKATCT 2216
Qy 2221 THGKAYDVDMVDANLF----MGGDVTRIESGSKVVVLDSLDPMVEERSDLEPSIPSEYML 2276
Db 2217 TRHDSPDADLIEANLLWRQEMGGNITRVESENKVVILDSFEPLQAEEDERGVSVPAEILR 2276
Qy 2277 PKKRFPPALPAWARPDYNPPLVESWKRPDYQPATVAGCALPPPRKTPTPPPRRRRTVGLS 2336
Db 2277 RSRKFPRAMPIWARPDYNPPLLESWKDPDYVPPVVHGCPLPPAKAPPIPPSRRKRTVVLS 2336
Qy 2337 EDSIGDALQQLAIKSFGQPPPSG-DSGLSTGAGAADSGSQTPPDELALSETGSISSMPPL 2395
Db 2337 ESTVSSALAELATETFGSSESSAVDSGTATA-----SPDQPSDDGDAGSDVESYSSMPPL 2391
Qy 2396 EGELGDPDLEPEQVEPQPPPQGGVAAPGSDSGSWSTCSEE-DDSVVCCSMSYSWTGALIT 2454
Db 2392 EGEPGDPDL----------SDGSWSTVSEEASEDVVCCSMSYTWTGALIT 2431
Qy 2455 PCSPEEEKLPINPLSNSLLRYHNKVYCTTTKSASLRAKKVTFDRMQVLDSYYDSVLKDIK 2514
Db 2432 PCAAEETKLPINALSNSLLRHHNLVYATTSRSANLRQKKVTFDRLQVLDDHYRDVLKEMK 2491
Qy 2515 LAASKVTARLLTMEEACQLTPPHSARSKYGFGAKEVRSLSGRAVNHIKSVWKDLLEDSET 2574
Db 2492 AKASTVKAKLLSVEEACKLTPPHSARSKFGYGAKDVRNLSSKAVNHIRSVWKDLLEDTET 2551
Qy 2575 PIPTTIMAKNEVFCVDPTKGGKKAARLIVYPDLGVRVCEKMALYDITQKLPQAVMGASYG 2634
Db 2552 PIDTTIMAKNEVFCVQPEKGGRKPARLIVFPDLGVRVCEKMALYDVVSTLPQAVMGSSYG 2611
Qy 2635 FQYSPAQRVEFLLKAWAEKKDPMGFSYDTRCFDSTVTERDIRTEESIYRACSLPEEAHTA 2694
Db 2612 FQYSPGQRVEFLVNAWKAKKCPMGFAYDTRCFDSTVTENDIRVEESIYQCCDLAPEARQA 2671

```
Qy       2695 IHSLTERLYVGGPMFNSKGQTCGYRRCRASGVLTTSMGNTITCYVKALAACKAAGIIAPT 2754
              | ||||||:|||: ||||| ||||||||||||||| |||:|||:|| |||:|| :   |
Db       2672 IRSLTERLYIGGPLTNSKGQNCGYRRCRASGVLTTSCGNTLTCYLKAAAACRAAKLQDCT 2731

Qy       2755 MLVCGDDLVVISESQGTEEDERNLRAFTEAMTRYSAPPGDPPRPEYDLELITSCSSNVSV 2814
              |||||||||| || ||:||| :||||||||||||||||:||||||||||||||||
Db       2732 MLVCGDDLVVICESAGTQEDEASLRAFTEAMTRYSAPPGDPPKPEYDLELITSCSSNVSV 2791

Qy       2815 ALGPQGRRRYYLTRDPTTPIARAAWETVRHSPVNSWLGNIIQYAPTIWARMVLMTHFFSI 2874
              |    |:| ||||||||||:||||||| ||:|||||||||| ||||:||||:||||||||
Db       2792 AHDASGKRVYYLTRDPTTPLARAAWETARHTPVNSWLGNIIMYAPTLWARMILMTHFFSI 2851

Qy       2875 LMAQDTLDQNLNFEMYGAVYSVSPLDLPAIIERLHGLDAFSLHTYTPHELTRVASALRKL 2934
              |:||: |:: |: ::||| ||: ||||| ||:||||| |||||:|:| |: |||| ||||
Db       2852 LLAQEQLEKALDCQIYGACYSIEPLDLPQIIQRLHGLSAFSLHSYSPGEINRVASCLRKL 2911

Qy       2935 GAPPLRAWKSRARAVRASLISRGGRAAVCGRYLFNWAVKTKLKLTPLPEARLLDLSSWFT 2994
              | |||| |: |||:||| |:|:||||| ||:||||||:|||||||:| |  |||||||
Db       2912 GVPPLRVWRHRARSVRARLLSQGGRAATCGKYLFNWAVRTKLKLTPIPAASQLDLSSWFV 2971

Qy       2995 VGAGGGDIYHSVSRARPRLLLFGLLLLFVGVGLFLLPAR 3033
               |  |||||||:||||||  :: |||| ||||::||| |
Db       2972 AGYSGGDIYHSLSRARPRWFMWCLLLLSVGVGIYLLPNR 3010


Search completed: October 28, 2005, 16:07:10
Job time : 83 secs
```